



XXVIII REUNIÓN CIENTÍFICA DEL GRUPO ARGENTINO DE BIOESTADÍSTICA

FORTALECIENDO LA SOBERANÍA A TRAVÉS DEL
CONOCIMIENTO CIENTÍFICO

LIBRO DE RESÚMENES

FACULTAD DE CIENCIAS EXACTAS Y NATURALES
UNIVERSIDAD DE BUENOS AIRES
CIUDAD DE BUENOS AIRES
8-10 OCTUBRE 2024





Grupo Argentino de Bioestadística

XXVIII Reunión científica del grupo Argentino de bioestadística. - 1a ed. - Neuquén : Grupo Argentino Bioestadística, 2025.

Libro digital, PDF

Archivo Digital: descarga

ISBN 978-631-90133-2-0

1. Bioestadísticas. 2. Estadísticas. 3. Análisis de Datos.

CDD 519.5

Esta es una obra colectiva. Derechos Reservados.

Diseño y edición: Valentín Rozenblit, Martín Echeverría y Adriana Pérez

ISBN 978-631-90133-2-0





Presentación

El Grupo Argentino de Bioestadística es una asociación que agrupa a profesionales de diversas instituciones públicas y privadas que se desempeñan como docentes, investigadores o consultores. Su misión es promover la investigación, el desarrollo, la aplicación, la enseñanza y la difusión de la teoría y los métodos estadísticos en las biociencias, abarcando disciplinas como biología, agronomía, medicina y veterinaria, entre otras. Asimismo, el Grupo Argentino de Bioestadística constituye la Región Argentina de la International Biometric Society (IBS, <https://www.biometricsociety.org>), una institución fundada en 1947 que reúne a investigadores en biometría de todo el mundo y está organizada en regiones geográficas para fomentar la colaboración científica internacional.

En este marco, entre los días 8 y 10 de octubre de 2024, se celebró la XXVIII Reunión Científica del Grupo Argentino de Bioestadística en el edificio 0+Infinito de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires, en la ciudad de Buenos Aires. Este evento tuvo como principal objetivo generar un espacio propicio para el intercambio y la discusión de ideas innovadoras dentro de la comunidad de especialistas en bioestadística. Se buscó divulgar los resultados de investigaciones recientes, fomentar la creación y el fortalecimiento de vínculos interinstitucionales y contribuir a la formación de recursos humanos en el área, promoviendo un entorno inclusivo y accesible para todos los participantes.

Este evento representó una valiosa oportunidad para explorar y compartir desarrollos y aplicaciones en el campo de la bioestadística y la ciencia de datos. La reunión contó con la participación de expertos nacionales e internacionales, quienes enriquecieron el evento a través de diversas actividades académicas. Entre ellas se destacaron conferencias, talleres, cursos breves, mesas de discusión y presentaciones de trabajos. Además, se llevó a cabo el tradicional concurso de Jóvenes Biometristas, en el cual estudiantes de grado y posgrado fueron desafiados a analizar un conjunto de datos y proponer una estrategia analítica. Los preseleccionados defendieron sus propuestas durante el evento, promoviendo el pensamiento crítico y la formación de nuevas generaciones en el campo.

El impacto de la XXVIII Reunión Científica del Grupo Argentino de Bioestadística se reflejó en una destacada participación y en la diversidad de actividades realizadas. El evento contó con 201 asistentes, de los cuales el 55% provenía del Área Metropolitana de Buenos Aires, mientras que el resto correspondía a participantes de diferentes provincias argentinas, así como de Uruguay y Chile. Cabe destacar que el 43% de los asistentes fueron estudiantes de



grado o doctorado, lo que resalta el compromiso de la reunión con la formación en bioestadística. Durante la reunión, se presentaron 107 resúmenes, distribuidos en 87 en formato póster y 20 en formato oral. Además, se llevaron a cabo 5 minicursos pre-congreso, que contaron con la asistencia de 142 personas. El programa académico incluyó 8 conferencias, 6 talleres y 2 mesas redondas, enriquecidas con la participación de 8 disertantes internacionales y 14 nacionales. En el Concurso de Jóvenes Biometristas, se presentaron 9 equipos de grado (14 participantes) y 6 equipos de posgrado (10 participantes), quienes demostraron su capacidad analítica y metodológica en la resolución de problemas aplicados. Asimismo, gracias al apoyo de auspiciantes, se otorgaron 25 becas destinadas a cubrir costos de inscripción y/o pasajes, favoreciendo así la participación de estudiantes e investigadores de distintos puntos del país.

Esta edición reafirma el compromiso del Grupo Argentino de Bioestadística con la excelencia académica y el fortalecimiento de la comunidad científica dedicada a la estadística, ciencias de datos e informática, convencidos del aporte de este conocimiento a las políticas basadas en evidencia. En consonancia con el lema de esta reunión, “Fortaleciendo la Soberanía a través del Conocimiento Científico”, reafirmamos que la inversión en ciencia y educación es fundamental para el progreso, la autonomía y el desarrollo sostenible de un país. Asimismo, reafirmamos la necesidad de garantizar la equidad y la inclusión en la producción y acceso al conocimiento, promoviendo espacios de intercambio abiertos, colaborativos y diversos.

Agradecemos a todos los participantes, disertantes y colaboradores por haber hecho posible este encuentro y esperamos que los contenidos compartidos en este libro sirvan como una fuente de inspiración y referencia para futuras investigaciones y desarrollos en el campo de la Bioestadística.

Comisión organizadora



Comisión Organizadora

Presidenta

Adriana Pérez - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Equipo

Teresa Boca - Facultad de Agronomía, Universidad de Buenos Aires

Pablo Cipriotti - Facultad de Agronomía, Universidad de Buenos Aires

Gerardo Cueto - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Magdalena Damiani Quiroz - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

María Soledad Fernández - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires - CONICET

Nicolás Flaibani - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires - CONICET

Tomás González Garello - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires - CONICET

Santiago Guidobono - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Elisa Helman - Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata - CONICET

Mathias Lopez - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Pablo Nuñez - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires - CONICET

Valentín Rozenblit - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires - CONICET

Marina Valdora - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Colaboradores

Ailen Altamirano - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Martín Echeverría - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Melisa Gutiérrez - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Solange Montilla - Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires



Comité Científico

Nora Abbiati - Universidad Nacional de Lomas de Zamora
Mónica Balzarini - Universidad Nacional de Córdoba
William Batista - Universidad de Buenos Aires
Sergio Bramardi - Universidad Nacional del Comahue
Cecilia Bruno - Universidad Nacional de Córdoba
Laura Calfayan - Universidad Nacional de San Martín
María Gabriela Cendoya - Universidad Nacional de Mar del Plata
Mariano Córdoba - Universidad Nacional de Córdoba
José Crespo - Universidad de Buenos Aires
Gerardo Cueto - Universidad de Buenos Aires
Pablo Demetrio - Universidad Nacional de La Plata
Julio Di Rienzo - Universidad Nacional de Córdoba
Julieta Filloy - Universidad de Buenos Aires
Adelia González Arzac - Universidad de Buenos Aires
Martín Graziano - Universidad de Buenos Aires
Valeria Gogni - Universidad de Buenos Aires
Pablo Inchausti - Universidad de la República
Diana Kelmansky - Universidad de Buenos Aires
Andrea Lavalle - Universidad Nacional del Comahue
María Llames - Universidad Nacional de San Martín
Joaquín Antonio Llera - Universidad Nacional de Cuyo
Javier Mariani - Universidad de Buenos Aires
Claudia Marinelli - Universidad Nacional del Centro de la Provincia de Buenos Aires
Fernando Milesi - Universidad Nacional del Comahue
Gloria Monterubbianesi - Universidad Nacional de Mar del Plata
Vanina Pagotto - Universidad Hospital Italiano de Buenos Aires
Susana Perelman - Universidad de Buenos Aires
Adriana Pérez - Universidad de Buenos Aires
Silvana Santellan - Universidad Nacional del Litoral
Silvia Sühling - Universidad Nacional de Salta



Comisión Directiva 2024-2026

Grupo Argentino de Bioestadística

Presidenta: Silvia Sühling. Cátedra de Estadística y Diseño Experimental, Facultad de Ciencias Naturales, Universidad Nacional de Salta

Secretaria: Verónica Lac Prugent. Cátedra de Estadística, Universidad Nacional del Comahue

Tesorera: Teresa Boca. Facultad de Agronomía, Universidad de Buenos Aires

Vocal 1: Cecilia Bruno. Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba. Unidad de Fisiología y Modelización Agrícola (UFyMA), CONICET

Vocal 2: Pablo Demetrio. Centro de Investigaciones del Medio Ambiente, CONICET – Universidad Nacional de La Plata

Vocal 1 Suplente: Mariano Córdoba. Estadística y Biometría. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba. Unidad de Fisiología y Modelización Agrícola (UFyMA), CONICET

Vocal 2 Suplente: Adriana Pérez. Grupo de Bioestadística Aplicada, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Revisora de Cuentas: Griselda Bóveda. Facultad de Ciencias Agrarias, Universidad Nacional del Nordeste

Revisora de cuentas suplente: Andrea Peña Malavera. Instituto de Tecnología Agroindustrial del Noroeste Argentino (ITANOA), Estación Experimental Agroindustrial Obispo Colombres (EEAOC).



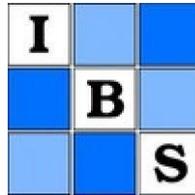
Auspiciantes



instituto de cálculo
UBA - CONICET



FAUBA



Grupo de Bioestadística Aplicada

CONICET



Instituto Nacional de
Tecnología Agropecuaria



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

GeoAgris®



NAVURE



Legislatura
Ciudad Autónoma de Buenos Aires



Índice

Presentación	3
Comisión Organizadora	5
Comité Científico	6
Comisión Directiva 2024-2026 Grupo Argentino de Bioestadística	7
Auspiciantes	8
Índice	9
Conferencias	16
Inferencia causal y Markoviana en la predicción de caracteres complejos empleando datos, ancestría e información genómica	17
NAVURE Software Estadístico: Novedades	17
¿Demasiadas variables? Una perspectiva general de la reducción de la dimensionalidad	18
Can Predictive machine learning models be used for statistical inference?	18
Los errores más comunes al analizar datos biológicos	19
Modelos de Regresión Poisson con inflación de ceros y datos censurados: predicción de éxito académico en cursos introductorios de matemática universitaria	19
Modelar múltiples variables de respuesta conjuntamente: boosting para la regresión distribucional y la regresión copular de distribuciones multivariantes	20
Estadística y ciencia de datos para la predicción: una larga saga de éxitos, yerros y desafíos	21
Talleres	22
¿Se puede inferir causalidad a partir de estudios observacionales?	23
Deconstruir el modelado no eStan complicado: una introducción a Stan	23
Análisis de supervivencia en presencia de riesgos competitivos	24
Regresión ordinal y multinomial	24
Training and Interpreting Deep Neural Networks with the cito R Package	24
Estadística frecuentista y bayesiana: dos miradas de un mundo regido por la sensibilidad	25
Variables predictoras co-lineales: Ridge y Lasso como métodos alternativos de selección	26
Mesas redondas	27
Mesa redonda I: Calidad de datos	28
Mesa redonda II: Educación estadística	28
Trabajos Presentados	30
Análisis Exploratorio y Visualización de datos	31
ANÁLISIS DE LAS INTERACCIONES HOSPEDADOR-PARÁSITO EN PECES DEL RÍO PARANÁ A TRAVÉS DE REDES BIPARTITAS	32
ANÁLISIS ESTADÍSTICO DE LA DINÁMICA DE LA INFECCIÓN POR VIRUS SINCICIAL RESPIRATORIO (VSR) Y ESTIMACIÓN DE SU INCIDENCIA EN EL AMBA	33
DATOS Y ARTE: EN UN ENFOQUE INNOVADOR	34
DESCRIPCIÓN Y ANÁLISIS DE LAS POBLACIONES Y COMUNIDADES DE HELMINTOS DE	



LAGOSTOMUS MAXIMUS (RODENTIA, CHINCHILLIDAE)	35
ESTADÍSTICA DESCRIPTIVA APLICADA AL MANEJO DE APIARIOS DE LA PROVINCIA DE ENTRE RÍOS	36
HIPERGRAFOS: UN NUEVO ENFOQUE METODOLÓGICO PARA DETERMINAR ESPECIES CLAVE EN UN ECOSISTEMA	37
Análisis Multivariado	38
COMPARACIÓN DE ALGORITMOS NIPALS Y SVD PARA REGRESIÓN POR MÍNIMOS CUADRADOS PARCIALES EN GWAS	39
ENTRENAMIENTO COGNITIVO DE FUNCIONES EJECUTIVAS EN NIÑOS ARGENTINOS: APLICACIÓN DE MÉTODOS DE CLUSTERING PARA EL ANÁLISIS DE PERFILES DE RENDIMIENTO	40
EVALUACIÓN DE INDICADORES METABÓLICOS EN CORDEROS TEXEL EMPLEANDO MODELOS LINEALES MIXTOS	41
PROTOCOLO DE ZONIFICACIÓN Y CARACTERIZACIÓN EDAFOCLIMÁTICA. APLICACIÓN EN UN VALLE VITIVINÍCOLA	42
ANÁLISIS DE REDUNDANCIA CON PARTICIÓN DE LA VARIANZA: UNA HERRAMIENTA ÚTIL PARA ESTUDIAR LA INFLUENCIA DE MÁS DE UN CONJUNTO DE VARIABLES AMBIENTALES EN LA VARIABILIDAD DE UNA COMUNIDAD DE AVES	43
PATRONES DE EXPRESIÓN EN PROTEÍNAS DE LA FAMILIA HSP20 EN ESTRÉS ABIÓTICO REVELADOS POR ANÁLISIS FACTORIAL MÚLTIPLE JERÁRQUICO	44
ANÁLISIS DE SERIES TEMPORALES PARA ESTUDIAR LA RELACIÓN ENTRE LA SEQUÍA Y LA VEGETACIÓN EN LOS BOSQUES SUBTROPICALES DE MISIONES	45
CALIDAD DE LECHE POR GRUPOS ETÁREOS EN BOVINOS CRIOLLOS ARGENTINOS	46
USO DE COVARIABLES CLIMÁTICAS Y EDÁFICAS PARA LA DEFINICIÓN DE “AMBIENTE” EN ESTUDIOS DE ESTABILIDAD DEL RENDIMIENTO DE MAÍZ EN ARGENTINA	47
DETECCIÓN DE ASOCIACIONES GENÉTICAS EN ACCESIONES DE QUÍNOA DEL NOROESTE ARGENTINO MEDIANTE MAPEO ASOCIATIVO CON VARIABLE DEPENDIENTE CUALITATIVA	48
CLASIFICACIÓN DE LA CALIDAD DEL HÁBITAT EN CURSOS DE AGUA DE LA PROVINCIA DEL NEUQUÉN A PARTIR DEL ÍNDICE IVH	49
DESARROLLO Y EVALUACIÓN DE UNA METODOLOGÍA DUAL PARA DATOS CATEGÓRICOS BASADA EN EL ANÁLISIS FACTORIAL MÚLTIPLE	50
APORTES DEL ANÁLISIS FACTORIAL MÚLTIPLE DUAL MIXTO EN EL ESTUDIO DE ESTRUCTURAS DE DATOS MULTIVÍA	51
EVALUACIÓN DE BIOESTIMULANTE A BASE DE ALGAS EN PLANTINES DE ALHELÍ (MATTHIOLA INCANA (L.) R.BR.)	52
ANÁLISIS MULTIVARIADO PARA LA EVALUACION DE FUNGICIDAS DE USO AGRÍCOLA E IDENTIFICACIÓN DE NUEVOS COMPUESTOS	53
ANÁLISIS MULTIVARIADO PARA LA EVALUACIÓN DE LA FUSIBILIDAD DE LAS CENIZAS DE BAGAZO Y RAC DE CAÑA DE AZÚCAR DE TUCUMÁN	54
ESTUDIO EXPLORATORIO SOBRE LAS PREFERENCIAS DEL CONSUMIDOR DE CARNE DE CORDERO	55
COMPARACIÓN DE LA PERFORMANCE DE MODELOS MORFOMÉTRICOS PREDICTORES	



DEL SEXO EN POMACEA (CAENOGASTROPODA: AMPULLARIIDAE)	56
AGRUPAMIENTOS BASADOS EN MIXTURAS FINITAS GAUSSIANAS: IMPACTO DE LA DIMENSIONALIDAD Y EL TAMAÑO DE MUESTRA EN DATOS SIMULADOS DE MORFOMETRÍA GEOMÉTRICA	57
ANÁLISIS DE MICROBIOMA RUMINAL DE CORDEROS TEXEL EN ENGORDE A CORRAL: EFECTO DE LA INCLUSIÓN DE LINO Y DEL TIPO DE MUESTREO RUMINAL	58
EVALUACIÓN DE DETERMINANTES DE LA MULTIFUNCIONALIDAD ECOSISTÉMICA DEL SUELO POR MEDIO DE SEM-PLS	59
Aprendizaje Automático	60
ENFOQUE DE APRENDIZAJE AUTOMÁTICO PARA LA SELECCIÓN DE VARIABLES APLICADO A XENOMETABOLOMICA	61
ESTADO DE AVANCE DEL ENSAYO PILOTO DE GENOMAS DE ALTA Y BAJA COBERTURA DEL PROGRAMA POBLAR	62
ALGORITMOS DE APRENDIZAJE AUTOMÁTICO EN LA PREDICCIÓN DEL COMPORTAMIENTO A ENFERMEDADES Y RENDIMIENTO EN DOS CULTIVOS	63
APLICACIÓN DEL DESCRIPTOR ESTADÍSTICO LBP PARA EL RECONOCIMIENTO AUTOMÁTICO DE TORTUGAS VERDES EN IMÁGENES	64
PREDICCIÓN DE DÍAS A MADUREZ EN SOJA (GLYCINE MAX SP.) A TRAVÉS DE IMÁGENES DE VEHÍCULOS AÉREOS NO TRIPULADOS	65
ELABORACIÓN DE UN MODELO DE PRONÓSTICO PARA FLORACIONES ALGALES EN EL EMBALSE SAN ROQUE, CÓRDOBA	66
DETECCIÓN DE CRIADEROS POTENCIALES DE AEDES ALBIFASCIATUS EN UNA CUENCA HÍDRICA METROPOLITANA, A PARTIR DEL ANÁLISIS DE IMÁGENES SATELITALES	67
DETECCIÓN DE DIFERENCIABILIDAD EN LA VOZ UTILIZANDO TÉCNICAS DE RESAMPLING SOBRE AUDIOS GRABADOS CON SMARTPHONES	68
MORFOMETRÍA GEOMÉTRICA Y REDES NEURONALES CONVOLUCIONALES: OPTIMIZACIÓN DE LA IDENTIFICACIÓN DE SUBESPECIES DE APIS MELLIFERA	69
CHAT GPT: ALIADO O ENEMIGO PARA EL ANÁLISIS ESTADÍSTICO DE DATOS	70
PREDICCIÓN DE RENDIMIENTOS AGRÍCOLAS EN CORRIENTES: INTEGRACIÓN DE DATOS MEDIANTE MODELOS DE APRENDIZAJE AUTOMÁTICO	71
¿CÓMO PREDIGO EL FUTURO?	
COMPARACIÓN ENTRE APRENDIZAJE AUTOMÁTICO Y PROFUNDO PARA LA PREDICCIÓN DE SERIES TEMPORALES	72
CIENCIA DE DATOS PARA SABER SI LOS MEDICAMENTOS PUEDEN LLEGAR AL CEREBRO	73
IMPACTO DEL TAMAÑO DE LA MUESTRA EN LA PRECISION DE MODELOS DE INTERPOLACION ESPACIAL. APLICACIONES EN AGRICULTURA	74
REDES NEURONALES APLICADAS EN LA PREDICCIÓN DE LA DEMANDA QUÍMICA DE OXÍGENO QUE INGRESA A LA PLANTA DE TRATAMIENTO DE LÍQUIDO CLOACAL	75
Bases de Datos y Datos Masivos	76
ESTIMACIÓN DE CORRELACIONES ENTRE RENDIMIENTOS DE CULTIVOS PROVENIENTES DE REGISTROS EMPRESARIALES	77
ESTUDIO DE ASOCIACIÓN DE GENOMA COMPLETO PARA MIOCARDIOPATÍA	



CHAGÁSICA EN LA PROVINCIA DE CÓRDOBA: RESULTADOS PRELIMINARES	78
CORRELACIONES ENTRE RENDIMIENTOS DE CULTIVOS PAMPEANOS ESTIMADAS A PARTIR DE ESTADÍSTICAS PÚBLICAS	79
Datos Espaciales	80
MODELANDO LA RESPUESTA DEL CULTIVO A LA APLICACIÓN DE INSUMOS POR ZONAS DE MANEJO EN ENSAYOS A ESCALA DE LOTE	81
ANÁLISIS DE PATRONES DE PUNTOS PARA DETECTAR ASOCIACIONES ESPACIALES ENTRE HORMIGAS DE DESIERTO Y CARACTERÍSTICAS DE SU ENTORNO	82
ACUICULTURA 4.0: ESTIMACIÓN BIODINÁMICA DE LA PRODUCCIÓN DE MEJILLONES	83
DETECCIÓN DE CAMBIOS EN LA DISTRIBUCIÓN ESPACIAL DE UNA COMUNIDAD MARINA MEDIANTE MODELOS NULOS	84
DESIGUALDADES TERRITORIALES EN LA REALIZACIÓN DE MAMOGRAFÍAS: ANÁLISIS A PARTIR DE MODELOS BAYESIANOS	85
ESTUDIO DE LA VARIABILIDAD TEMPORAL DE PASTIZALES INVADIDOS USANDO SERIES TEMPORALES DE IMÁGENES SENTINEL-2	86
INTERPOLACIÓN ESPACIAL DE MATERIA ORGÁNICA DEL SUELO A ESCALA DE LOTE USANDO QUANTILE REGRESSION FOREST	87
COMPARACIÓN DE MÉTODOS DE VALIDACIÓN CRUZADA EN SERIES DE TIEMPO SIMULADAS UTILIZANDO MODELO ARIMA PARA PRONÓSTICO DE PRECIO	88
ANÁLISIS DE LA VARIABILIDAD ESPACIO-TEMPORAL DEL RENDIMIENTO EN UN SITIO CON VID VAR. CABERNET SAUVIGNON PARA PRONÓSTICOS DE PRODUCCIÓN Y ZONAS CON CALIDADES ENOLÓGICAS DIFERENCIADAS	89
ANÁLISIS DE LAS VARIACIONES TEMPORALES EN LAS BRECHAS DE NIVEL SOCIOECONÓMICO Y GÉNERO EN EL CONSUMO DE FRUTAS Y VERDURAS EN LA POBLACIÓN ADULTA DE ARGENTINA A PARTIR DE DATOS PÚBLICOS	90
DIVERSIDAD- β Y REGISTRO FÓSIL, MITIGACIÓN DE PSEUDOPATRONES	91
MODELOS GEOESTADÍSTICOS Y DE SERIES DE TIEMPO PARA ANALIZAR LA CALIDAD DEL AIRE EN LA CIUDAD DE SAN SALVADOR DE JUJUY	92
RIESGO RELATIVO DE INFECCIONES RESPIRATORIAS Y VIRUS SINCICIAL EN EL AMBA: UNA MIRADA DESDE LA ESTADÍSTICA ESPACIAL PARA ANALIZAR SU DINÁMICA PRE Y POSTPANDEMIA	94
APLICACIÓN DE GAMM EN EL ESTUDIO DE LA RESPUESTA INMUNE HUMORAL POST VACUNACIÓN CONTRA COVID-19 EN ADULTOS MAYORES INSTITUCIONALIZADOS	95
ANÁLISIS DE MODELOS JERÁRQUICOS EN DATOS DE ESTRUCTURA ANIDADA UTILIZANDO DISTRIBUCIONES POISSON GENERALIZADAS	96
Datos Longitudinales y Series Temporales	97
‘¿CÓMO HARÍAS PARA SABER CUÁNTO PIS HACE UNA PERSONA EN UN DÍA?’ DESARROLLO DE HABILIDADES DE PENSAMIENTO CIENTÍFICO EN ADOLESCENTES ARGENTINOS	98
ANÁLISIS DE LOS EFECTOS DEL HIPOTIROIDISMO EN DIFERENTES ETAPAS DE LA VIDA SOBRE EL CRECIMIENTO Y PERFIL HORMONAL	99
SALUD DE LAS COLONIAS DE ABEJAS EN SISTEMAS APÍCOLAS SANTAFESINOS DE	



PEQUEÑA ESCALA	100
MODELOS LINEALES GENERALIZADOS MIXTOS PARA COMPARAR ANTIPARASITARIOS DESDE EL CONTEO DE NEMATODOS EN OVEJAS CRIOLLAS	101
MÉTRICAS FENOLÓGICAS DEL BOSQUE DE LENGAS EN EL NO DE CHUBUT UTILIZANDO DIFERENTES ESCALAS TEMPORALES DE DATOS DE NDVI-MODIS	102
Educación estadística	103
INTEGRACIÓN DEL SOFTWARE R, RSTUDIO Y POSIT CLOUD EN LA EDUCACIÓN ESTADÍSTICA: EXPERIENCIA CON ESTUDIANTES DE	104
INNOVACIÓN EN LA ENSEÑANZA DE ESTADÍSTICA EN CIENCIAS AGRARIAS: INTEGRACIÓN DE R Y RSTUDIO EN ENTORNOS EDUCATIVOS	105
ANÁLISIS DE LOS SABERES PREVIOS DE LOS ESTUDIANTES DE BIOESTADÍSTICA	106
EVALUACIÓN DE LOS CONOCIMIENTOS PREVIOS DE LOS ESTUDIANTES DE ÓPTICA GEOMÉTRICA LUEGO DEL CAMBIO DE PLAN DE ESTUDIOS	107
VISUALIZACIÓN DE LA SITUACIÓN ARGENTINA A PARTIR DE INDICADORES DE ACCESO PÚBLICO	108
DISTRIBUCIÓN GEOGRÁFICA Y ANÁLISIS SOCIOECONÓMICO DE LOS INGRESANTES A LA FACULTAD DE ECONOMÍA Y ADMINISTRACIÓN (FAEA) DE LA UNIVERSIDAD NACIONAL DEL COMAHUE, 2024	109
APRENDIZAJE AUTODIRIGIDO: CLAVE PARA EL DESARROLLO DE HABILIDADES DE PENSAMIENTO ESTADÍSTICO	110
CUALIDADES DE LOS INSTRUMENTOS PARA LA EVALUACIÓN SUMATIVA DE CONCEPTOS ESTADÍSTICOS APLICADOS EN PROBLEMAS DE CIENCIAS AGROPECUARIAS	111
Meta-Análisis	112
META-ANÁLISIS COMO HERRAMIENTA PARA LA EVALUACIÓN DE RESPUESTAS DE AVES AL MANEJO GANADERO: UN ESTUDIO DE CASO EN LOS PASTIZALES DEL RÍO DE LA PLATA	113
DESAFÍOS METODOLÓGICOS EN UNA REVISIÓN SISTEMÁTICA VIVA Y META-ANÁLISIS INTERACTIVOS QUE EVALÚA LA SEGURIDAD Y EFICACIA DE VACUNAS COVID 19 EN EMBARAZADAS	114
Métodos Bayesianos	115
EFFECTO DE LOS MODELACIÓN DE LA PRESENCIA DEL CARBÓN DE LA ESPIGA DEL MAÍZ (<i>Ustilago maydis</i>) EN LA IDENTIFICACIÓN DE RESISTENCIA	116
VEROSIMILITUD PENALIZADA CON APORTES BAYESIANOS PARA ANALIZAR LA LOCALIZACIÓN DE TERMORRECEPTORES EN LAS ANTENAS DE RHODNIUS PROLIXUS	117
ESTIMACIÓN DE VARIANZA GENÉTICA EN LA RESISTENCIA A INSECTICIDAS	118
ESTIMACIÓN DE LA EDAD DE LOS ROEDORES USANDO MODELOS BAYESIANOS	119
Métodos Estadísticos	120
CÁLCULO DE UNA F DE COHEN EFICIENTE PARA UNA PRUEBA ÓMNI-BUS	121
Métodos Robustos y No Paramétricos	122
HERRAMIENTAS PARA EL ANÁLISIS ROBUSTO DE DATOS DOSIS-RESPUESTA	123



INTERPRETANDO LAS DESVIACIONES DE LA PREDICCIÓN ALOMÉTRICA EN MACROEVOLUCIÓN: UN ENFOQUE UTILIZANDO TASAS DE CAMBIO	124
Modelos Aditivos Generalizados	125
LA EVAPOTRANSPIRACIÓN EN BOSQUES MISIONEROS CON Y SIN EXTRACCIÓN SELECTIVA DE MADERA	126
DESIGUALDADES EN LA REALIZACIÓN DE ACTIVIDAD FÍSICA Y DE SEDENTARISMO EN ADULTOS SEGÚN CARACTERÍSTICAS CLIMÁTICAS Y SOCIODEMOGRÁFICAS EN ARGENTINA: UN ESTUDIO A PARTIR DE LA ENCUESTA NACIONAL DE FACTORES DE RIESGO 2018	127
Modelos de Simulación	128
MODELO ESTOCÁSTICO DE LA DINÁMICA DE PROPAGACIÓN DE PESTE PORCINA AFRICANA EN EL SECTOR PORCINO DE LA ARGENTINA	129
Modelos de Supervivencia	130
DEL EFECTO DEL RIESGO DE DEPREDACIÓN SOBRE LA LATENCIA DE RETORNO AL NIDO MEDIANTE MODELOS DE REGRESIÓN DE COX	131
SUPERVIVENCIA APARENTE Y TASA DE RETORNO DEL JILGUERO DORADO (SICALIS FLAVEOLA PELZELNI), UNA ESPECIE CON FENOTIPO RELACIONADOS CON LA EDAD	132
ECTOPARASITICIDAS APLICADOS A GATOS DOMÉSTICOS COMO HERRAMIENTA DE CONTROL DE T. INFESTANS RESISTENTES A LOS PIRETROIDES	133
Modelos Lineales Generales	134
ANÁLISIS DE VARIABLES NORMALES DISCRETIZADAS: UN CASO DE ESTUDIO	135
MODELO FACTORIAL DE ANOVA PARA LA BÚSQUEDA DE ANTIFÚNGICOS AMIGABLES CON EL MEDIO AMBIENTE Y SU APLICACIÓN EN EL PATRIMONIO CULTURAL	136
VIRULENCIA PARASITARIA EN RATAS COINFECTADAS: COMPARACIÓN DE DOS MÉTODOS ANALÍTICOS	137
EFFECTOS DE LA SOBREALIMENTACIÓN POSTNATAL TEMPRANA EN RATAS ADULTAS, SU MODELADO A TRAVÉS DE UN DISEÑO DE MEDIDAS REPETIDAS	138
RENDIMIENTO Y ESTABILIDAD DE HÍBRIDOS Y VARIEDADES DE COLZA PRIMAVERAL EN MÚLTIPLES AMBIENTES	139
METODO STATICO: LA IMPORTANCIA DEL AGUA DULCE: MONITOREO DE ARROYOS Y RÍOS CORDILLERANOS DE NEUQUEN	140
Modelos Lineales Generalizados	141
EXTRACCIÓN DE HONGOS EN PEDICULUS HUMANUS CAPITIS	142
DIFERENCIACIÓN DE VARIEDADES DE CANNABIS SATIVA SEGÚN LA COMPOSICIÓN DE CANNABINOIDES EMPLEANDO ANÁLISIS MULTIVARIADO	143
ESTIMACIÓN NO DESTRUCTIVA DE ÁREA FOLIAR EN CLONES DE CACAO	144
ESTUDIO COMPARATIVO DE TÉCNICAS ANALÍTICAS PARA LA EPIDEMIOLOGÍA VETERINARIA	145
APLICACIÓN DE MODELOS DE REGRESIÓN LOGÍSTICA PARA EVALUAR EL IMPACTO DE DIFERENTES FACTORES SOBRE LA TASA DE PREÑEZ EN VACAS	146
MODELOS DE REGRESIÓN LOGÍSTICA PARA PREDECIR LA HABILIDAD DE	



PERMANENCIA EN HEMBRAS BOVINAS CRIOLLAS	147
INTERACCIÓN ENTRE FACTORES AMBIENTALES PARA ENTENDER LA RESPUESTA DE LA RIQUEZA DE AVES A GRADIENTES URBANO-RURALES	148
DISFUNCIONES DEL PISO PÉLVICO EN MUJERES ARGENTINAS QUE REALIZAN ACTIVIDAD FÍSICA Y SU IMPACTO FUNCIONAL. ESTUDIO EPIDEMIOLÓGICO, ANALITICO	149
RESTAURACIÓN DE PASTIZALES INVADIDOS: EVALUACIÓN EXPERIMENTAL DE TÉCNICAS DE MANEJO CON DISEÑO FACTORIAL	150
CONSIDERACIÓN DEL DISEÑO DE MUESTREO EN EL ANÁLISIS ESTADÍSTICO DE PATRONES ALTITUDINALES DE RIQUEZA	151
INCIDENCIA DEL COEFICIENTE DE FRUTO OCULTO EN LAS ESTIMACIONES DE PRODUCCIÓN DE CÍTRICOS DE LA PROVINCIA DE CORRIENTES	152
IMPACTO EN LA ESTIMACIÓN DE LA MAGNITUD DEL EFECTO Y EN LA PROBABILIDAD DE ERROR DE TIPO I, AL OMITIR LA ESTRUCTURA JERÁRQUICA EN UN MODELO LINEAL GENERALIZADO MIXTO	153
Modelos No Lineales	154
APLICACIÓN Y ANÁLISIS DE LAS CURVAS DE PERFORMANCE TÉRMICA EN EL ESTUDIO DEL NICHOS ECOLÓGICO DE ESPECIES DE DROSOPHILA	155
SELECCIÓN DE MODELOS DE CRECIMIENTO POBLACIONAL EN <i>Daphnia pulicaria</i> : UNA EVALUACIÓN MEDIANTE NLS Y AIC	156
SELECCIÓN DE PARÁMETROS PARA DESCRIBIR EL CRECIMIENTO EN FRUTOS DE MANDARINO TANGOR MEDIANTE MODELOS NO LINEALES MIXTOS	157
Trabajos Extendidos	158
ANÁLISIS ESTADÍSTICO DE LA DINÁMICA DE LA INFECCIÓN POR VIRUS SINCICIAL RESPIRATORIO Y ESTIMACIÓN DE SU INCIDENCIA EN EL AMBA	159
MODELANDO LA RESPUESTA DEL CULTIVO DE MAÍZ A LA APLICACIÓN DE INSUMOS POR ZONAS DE MANEJO EN ENSAYOS A ESCALA DE LOTE	164
ANÁLISIS DE PATRONES DE PUNTOS PARA DETECTAR ASOCIACIONES ESPACIALES ENTRE HORMIGAS DE DESIERTO Y CARACTERÍSTICAS DE SU ENTORNO	169
INTERPOLACIÓN ESPACIAL DE MATERIA ORGÁNICA DEL SUELO A ESCALA DE LOTE USANDO QUANTILE REGRESSION FOREST	186
USO DE COVARIABLES CLIMÁTICAS Y EDÁFICAS PARA LA DEFINICIÓN DE "AMBIENTE" EN ESTUDIOS DE ESTABILIDAD DEL RENDIMIENTO DE MAÍZ EN ARGENTINA	191
CLASIFICACIÓN DE LA CALIDAD DEL HÁBITAT EN CURSOS DE AGUA DE LA PROVINCIA DEL NEUQUÉN A PARTIR DEL ÍNDICE IVH	196
RIESGO RELATIVO DE INFECCIONES RESPIRATORIAS Y VIRUS SINCICIAL EN EL AMBA: UNA MIRADA DESDE LA ESTADÍSTICA ESPACIAL PARA ANALIZAR SU DINÁMICA PRE Y POSTPANDEMIA	200
Índice por autor	211



Conferencias



Inferencia causal y Markoviana en la predicción de caracteres complejos empleando datos, ancestría e información genómica

CANTET, RODOLFO J. C.

Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires

En esta charla se presenta la modelación estocástica de procesos Markovianos susceptibles de determinar el valor genético de un individuo para caracteres complejos como la producción de un cultivo o un árbol, el nivel productivo (leche, carne, lana, huevos) de un animal, o la susceptibilidad a una enfermedad con una componente hereditaria, cuando existe información de datos, ancestría e información genómica. La presentación discute el valor de la información de parientes con matrices de covarianzas que reflejan dependencia positiva multivariada total (MTP2) de procesos individuales como la recombinación, como alternativa a procesos poblacionales que sobreparameterizan el modelo predictivo generando excesivo error cuadrático medio de predicción. Finalmente, se presenta la predicción empleando matrices de covarianza que refleja procesos causales Markovianos, sobre la base de segmentos compartidos de identidad por descendencia, y su aplicación a la evaluación genética animal.

NAVURE Software Estadístico: Novedades

DI RIENZO, JULIO ALEJANDRO

Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba

El año 2024 ha sido un período de cambios importantes para NAVURE. Los cambios más significativos han sido a nivel de gráficos, reporte de resultados y el refinamiento de la interfaz del usuario en muchos procedimientos estadísticos. En los gráficos, se añadieron muchas opciones de personalización entre las que podemos citar la edición y relocalización de etiquetas, coloración individual de elementos y ajuste de opacidad por serie. Los gráficos de barras y de puntos se pueden alternar para visualizarse como gráficos de área y viceversa y aparecer combinadas. En los modelos de regresión lineal y lineal generalizada la salida incluye gráfico de residuos parciales y las fórmulas de regresión estimadas. En los gráficos se pueden etiquetar los puntos observados y se pueden añadir con facilidad potencias y productos de covariables y sus interacciones con factores para ajustar modelos polinómicos. La exploración y comparación de modelos se han separado, y las pruebas de combinaciones lineales ahora incluyen una prueba simultánea. En regresión no lineal, ahora se visualizan fórmulas explícitas. Se pueden añadir múltiples factores en efectos fijos o aleatorios, etiquetar puntos en los gráficos y se incorporaron colores para identificar los puntos observados según los niveles de los factores involucrados. Un aspecto destacado es la capacidad de cancelar procesos largos o ejecutarlos en segundo plano.

Ciudad de Buenos Aires, octubre 2024



En esta presentación repasaremos los cambios introducidos en el marco de la aplicación de modelos lineales generalizados mixtos en un problema de reproducción animal que incluye datos binarios y de conteo.

¿Demasiadas variables? Una perspectiva general de la reducción de la dimensionalidad

FORTE, ANABEL

*Departamento de Estadística e Investigación Operativa, Facultad de Ciencias Matemáticas,
Universidad de Valencia, Valencia, España*

Reducir la dimensión del modelo o el número de variables con el que trabajamos en un estudio estadístico es una práctica esencial en la investigación científica cuya aplicación puede determinar la precisión y validez de los resultados obtenidos. En cualquier estudio científico, la elección de las variables adecuadas no solo simplifica los modelos y mejora su interpretabilidad, sino que también reduce el riesgo de sobreajuste y permite que los resultados sean generalizables a situaciones del mundo real. A lo largo de esta conferencia repasaremos cuáles son las principales herramientas con las que contamos para realizar esta reducción de la dimensionalidad y la importancia de éstas para mejorar la calidad de los análisis científicos. Desde los métodos por pasos basados en p-valores o en criterios de información, a la perspectiva bayesiana de selección de variables basada en probabilidades a posteriori, pasando por las técnicas de reducción de la dimensionalidad que, si bien no seleccionan variables de forma individual, sí nos ayudan a encontrar patrones de comportamiento tanto en las variables como en los datos, basándonos en combinaciones de las variables originales. Además, en esta tarea no podemos dejar de lado la consideración de situaciones que suponen un reto como son la presencia de datos faltantes o la correlación entre las variables sobre las que se quiere realizar la selección o agrupación.

Can Predictive machine learning models be used for statistical inference?

HARTIG, FLORIAN

*Grupo de Ecología Teórica, Facultad de Biología y Medicina Pre-Clínica, Universidad de Regensburg,
Alemania*

The popularity of machine learning (ML), deep learning (DL), and artificial intelligence (AI) has grown rapidly in recent years. Although ML and DL are often thought to excel at prediction, they can, at least in principle, also be used for analytical tasks traditionally in the domain of statistical models, such as effect size estimation or partitioning of explained variation. In my talk, I will discuss the promise and pitfalls of this path. I will review the different methods proposed in the field of explainable AI (xAI) that could be used to extract effects and variable importance from fitted

Ciudad de Buenos Aires, octubre 2024



models, and how these map to traditional statistical indicators. As a next step, I will discuss how we could provide statistical guarantees such as p-values and confidence intervals for these xAI metrics. A key problem that arises in this context is understanding what inductive biases are introduced by the various ML algorithms and their hyperparameters, and how these affect the bias in xAI effect estimates. I will conclude that there is reason to be cautiously hopeful that ML algorithms may merge with other methods for statistical inference, which opens up an exciting new area for statistical research.

Los errores más comunes al analizar datos biológicos

IENO, ELENA

Highland Statistics Ltd.

La investigación científica tiene valor sólo en la medida en que se obtienen y comunican resultados relevantes, una tarea que demanda organizar, evaluar, analizar y comunicar sin ambigüedades la importancia de los datos. El análisis de datos ecológicos que reflejen las complejidades e interacciones del mundo natural constituye, sin lugar a duda, un desafío para la enseñanza de la estadística a nivel de grado y posgrado. Los progresos recientes para el análisis estadístico de datos multifacéticos hacen posible la obtención de resultados más precisos y significativos. Sin embargo, las decisiones claves de las herramientas a utilizar, la interpretación de los resultados y la validación de los modelos constituyen etapas fundamentales que, de no aplicarse, llevarían a conclusiones falsas y recomendaciones incorrectas. Sin lugar a dudas existen bases de datos que requieren distintas estrategias para la solución de problemas ya que en ciertas situaciones se hace complejo poder coincidir las hipótesis de trabajo con la estructura de datos de campo. De ahí la importancia de poder descubrir determinadas problemáticas tales como: el uso de covariables categóricas con varios niveles y un tamaño de muestra bajo (“Quasi complete separation/complete separation”), multicolinealidad de difícil detección (concurvity), inconsistencia de la resolución espacio-temporal, pseudoreplicación (múltiples variables de respuesta en el mismo modelo), entre otras. El objetivo de esta charla es proporcionar una descripción general de los errores estadísticos más frecuentes encontrados en los trabajos científicos y enumerar algunos consejos útiles sobre cómo y cuándo evitarlos.

Modelos de Regresión Poisson con inflación de ceros y datos censurados: predicción de éxito académico en cursos introductorios de matemática universitaria

MACHIAVELLI, RAÚL

Facultad de Ciencias Agrícolas, Departamento de Agronomía y Tierra, Universidad de Puerto Rico,

Ciudad de Buenos Aires, octubre 2024



Mayagüez, Puerto Rico

Los cursos de matemática a nivel universitario representan los “cuellos de botella” en las carreras universitarias que los requieren. Existen numerosas formas de relacionar el éxito en estos cursos con variables demográficas y variables que describen el rendimiento académico en el nivel secundario. Por ejemplo, podemos intentar predecir la cantidad de veces que un estudiante repite una clase hasta aprobarla usando variables que conocemos en el momento que el estudiante comienza sus estudios, y así diseñar actividades para fortalecer su aprendizaje. Es de esperar que al momento de recolectar los datos aun existan alumnos que no hayan aprobado la clase, por lo que desconocemos el número de veces que estos alumnos repetirán hasta aprobar, por lo que esta observación será censurada por la derecha (solo sabemos que esta cantidad es mayor que el valor observado). Por otro lado, un alto porcentaje de alumnos aprueba la materia sin repetirla, es decir, existe un gran porcentaje de ceros. En estos casos, tendremos datos inflados en cero con censura. En esta charla proponemos un modelo de regresión Poisson con inflación de ceros y datos censurados, para lo cual plantearemos ecuaciones de verosimilitud computacionalmente eficientes y estudiaremos algunas de sus propiedades usando simulación. Finalmente, aplicaremos este modelo a un conjunto de datos de estudiantes de la Universidad de Puerto Rico para encontrar modelos que predigan si un estudiante está en riesgo de reprobar las clases introductorias de matemáticas y cuántas veces repetiría la clase (en promedio) si no se realizan intervenciones académicas adicionales.

Modelar múltiples variables de respuesta conjuntamente: boosting para la regresión distribucional y la regresión copular de distribuciones multivariantes

MAYR, ANDREAS

Department of Medical Biometry and Statistics, University of Marburg, Germany

Los modelos de regresión tradicionales a menudo proporcionan una visión excesivamente simplista sobre las asociaciones y relaciones complejas en los problemas contemporáneos de datos en el área de la biomedicina. En particular, capturar correctamente las asociaciones relevantes entre múltiples situaciones clínicas finales es de gran importancia para evitar especificaciones incorrectas del modelo, lo que puede llevar a resultados sesgados e incluso a conclusiones erróneas o engañosas. Para modelar múltiples resultados clínicos conjuntamente, se puede recurrir a distribuciones multivariadas como una distribución Gaussiana o Binomial bivariada. En el contexto de la regresión distribucional, la idea es modelar no solo los parámetros de ubicación, sino todos los parámetros de la distribución multivariada de forma conjunta. Sin embargo, en algunos contextos necesitaremos aún más flexibilidad. En esta charla presentaré también nuevos modelos de regresión condicional con cópulas para estructuras de datos biomédicos de alta dimensión, combinando herramientas de aprendizaje estadístico eficiente para datos de alta dimensión con métodos establecidos en economía que permiten capturar estructuras de dependencia complejas. Estos métodos nos permitirán modelar toda la distribución conjunta de múltiples puntos finales

Ciudad de Buenos Aires, octubre 2024



simultáneamente y determinar automáticamente las covariables influyentes y los factores de riesgo relevantes mediante algoritmos de boosting propuestos originalmente en el área de estadística y aprendizaje automático. Ilustraré todos los modelos con ejemplos de investigaciones recientes de nuestro equipo, destacando también en qué contextos los nuevos enfoques aún presentan problemas o cuándo modelos más simples ofrecen un rendimiento similar

Estadística y ciencia de datos para la predicción: una larga saga de éxitos, yerros y desafíos

SOSA ESCUDERO, WALTER

Departamento de Economía, Universidad de San Andrés, Buenos Aires

De la meteorología a la economía, pasando por la religión, el fútbol y la política, esta conferencia discute el rol de los pronósticos científicos, por qué funcionan, por qué no y por qué ciertas cuestiones son esencialmente impredecibles. En un lenguaje informal, pero con el mayor rigor científico, esta conferencia revisa el rol de los datos, la estadística y la inteligencia artificial en cuestiones predictivas de diversas áreas del conocimiento.



Talleres



¿Se puede inferir causalidad a partir de estudios observacionales?

BABINO, LUCÍA

*Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires –
CONICET*

La mayoría de las veces que realizamos un análisis estadístico, estamos motivados por una pregunta causal. La regla de oro para responder preguntas causales son los experimentos controlados en los que el tratamiento se asigna aleatoriamente, todos los sujetos cumplen con el tratamiento y el outcome se mide sin error. En estos estudios, es correcto interpretar una medida de asociación como medida de efecto causal. Sin embargo, muchos experimentos distan de ser ideales. Más aún, para ciertas preguntas realizar experimentos es imposible, demasiado costoso o simplemente no es ético y los únicos datos disponibles son observacionales. Es en estas situaciones en las que la teoría de la Inferencia Causal resulta indispensable.

En este taller introduciremos algunos conceptos clave de la Inferencia Causal a través de un enfoque gráfico basado en DAGs. Veremos cómo los DAGs pueden ayudarnos a determinar si los datos disponibles son suficientes para responder nuestra pregunta causal y, en caso afirmativo, por qué variables debemos controlar en un modelo de regresión para poder estimar sin sesgo el efecto causal de interés

Deconstruir el modelado no eStan complicado: una introducción a Stan

BARBERÁ, IVÁN

Instituto de Investigaciones en Biodiversidad y Medioambiente (INIBIOMA), CONICET – Universidad Nacional del Comahue

Si la variable explicativa es categórica, ANOVA; si es cuantitativa, regresión; si no se cumple el supuesto de normalidad, GLM, y así hasta las ramas más recónditas de la clave dicotómica. A veces la estadística parece una bolsa de métodos inconexos acompañada de una receta para usarlos. Bajo este enfoque, elegimos qué modelo utilizar siguiendo la receta y evaluando la disponibilidad de software. Pero las herramientas computacionales actuales permiten adoptar otro enfoque: partiendo de nuestras preguntas u objetivos, podemos formular un modelo que represente el proceso bajo estudio de manera razonable, y luego buscar la forma de estimarlo. En este taller veremos cómo dar los primeros pasos hacia un modelado más libre de recetas usando Stan, una plataforma estadística orientada a la estimación de modelos estadísticos con enfoque Bayesiano.



Análisis de supervivencia en presencia de riesgos competitivos

CARBONELL-ASINS, JUAN ANTONIO

INCLIVA Instituto de Investigación Sanitaria, Universidad de Valencia, España

El objetivo de la charla es profundizar en los métodos de análisis de supervivencia. En esta ocasión nos centraremos en los modelos de riesgos competitivos que podrían verse como una extensión del modelo de regresión de Cox para estudiar un evento que solo puede ocurrir si un segundo evento no ocurrió previamente. Todo el seminario tendrá una perspectiva principalmente práctica en la que, utilizando ejemplos reales, discutiremos la utilidad práctica de estos modelos poniendo el énfasis en las diferencias y similitudes con el modelo de regresión de Cox. Finalmente veremos un caso real de riesgos competitivos en el contexto de machine learning.

Regresión ordinal y multinomial

INCHAUSTI, PABLO

Departamento de Ecología y Evolución, Universidad de la República, Uruguay

Los modelos lineales generalizados (GLM) permiten analizar muchos tipos de variables de respuesta (binaria, conteos, valores reales estrictamente positivos, etc) empleando un método común para la estimación de parámetros y la realización de inferencia. El marco teórico original de los GLM fue extendido para analizar variables de respuesta tales como proporciones, conteos con excesos de variación y de ceros, etc. Los GLM multinomiales y ordinales permiten analizar variables de respuesta politómicas (i.e. que tienen más de un estado posible, todos mutuamente excluyentes y que podrían o no conformar una jerarquía natural de valores) que se modelan con la función de distribución de probabilidades Multinomial. Los GLM multinomiales y ordinales son realmente una variación de los GLM binarios pero que emplean funciones de enlace logit particulares (nada muy complicado). Este taller se hará una introducción general a estos GLM, se mostrará su ajuste con paquetes de R y la interpretación de los parámetros estimados con ejemplos específicos, y se discutirán la evaluación de su calidad de ajuste a través de análisis de residuos particulares. Si bien los GLM multinomiales y ordinales no son particularmente nuevos y se emplean en medicina, economía y marketing, su uso es relativamente poco conocido en ciencias biológicas donde podrían permitir adquirir nuevos datos y analizar otros que serían generalmente descartados por desconocimiento de estos métodos analíticos.

Training and Interpreting Deep Neural Networks with the cito R Package

PICHLER, MAXIMILIAN

Ciudad de Buenos Aires, octubre 2024



Faculty of Biology and Pre-Clinical Medicine, University of Regensburg, Germany

cito is an easy-to-use R package for deep learning that allows users to build neural networks using the familiar formula syntax of many other R packages. Yet, *cito* allows flexible modification of network architecture and hyperparameters, allowing users to test and apply most deep learning techniques without having to focus on coding. In addition, *cito* includes many user-friendly functions for model interpretation based on explicable AI and additional features such as confidence intervals (p-values). In this workshop, we will first introduce *cito* and explain how to train deep neural networks (DNN) on tabular data. We will address common challenges associated with DNN training, such as convergence issues, and show how the model can be modified to optimize the bias-variance trade-off. We will also explain how to extract effects (similar to linear effects from regression models) and variable importance (similar to ANOVA) from the fitted DNN and how to interpret these effects. Finally, we will show how *cito* can be used to fit CNNs for image classification and regression.

Estadística frecuentista y bayesiana: dos miradas de un mundo regido por la sensibilidad

RISSO, PAULA

Facultad de Ciencias Veterinarias y Centro Universitario de Farmacología, Universidad Nacional de La Plata

La mayoría de las investigaciones biológicas se presentan con un enfoque analítico o inferencial, dándole un papel fundamental tanto al concepto de valor-p (o p-value) como a la decisión de rechazar (o no) a la hipótesis nula. Bajo esta premisa, resulta interesante desarrollar dos caminos posibles dentro del análisis de datos, que permitirían dar respuesta a cualquier problemática: la visión clásica o frecuentista y la perspectiva bayesiana. De esta manera, se podrá estudiar la sensibilidad de ambas metodologías. En este taller se introducirán conceptos básicos del teorema de Thomas Bayes, como distribución Prior (informativa y no informativa), distribución Posterior, Intervalo de Credibilidad Bayesiano (ICB) y el proceso inferencial involucrado. Se presentarán distintos casos problema, cuya resolución se verá desde la visión frecuentista y desde la perspectiva bayesiana, utilizando distribuciones Prior no informativas. A su vez, para analizar y comparar el comportamiento de ambos paradigmas, se realizarán simulaciones sobre el tamaño muestral mínimo necesario para encontrar diferencias significativas, en distintos escenarios posibles. Finalmente, se presentarán y discutirán las ventajas de cada mirada, en el presente y el futuro inmediato.

Ciudad de Buenos Aires, octubre 2024



Variables predictoras co-lineales: Ridge y Lasso como métodos alternativos de selección

SZRETTTER NOSTE, MARÍA EUGENIA

*Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires –
CONICET*

En diferentes disciplinas los estudios observacionales son frecuentes y la multicolinealidad es un problema al que se debe hacer frente al momento de modelar. Las técnicas de stepwise regression han sido altamente difundidas como métodos de selección de modelos. Pero... tienen sus limitaciones. En este taller presentaremos dos técnicas de regularización como propuestas de métodos alternativos para la selección de variables. Abordaremos algunas ideas principales sobre las técnicas y luego nos abocaremos a su implementación en R y posterior interpretación de los resultados.



Mesas redondas



Mesa redonda I: Calidad de datos

CECILIA RUZ¹, SABRINA LÓPEZ², AGUSTÍN DRAMIS³, PABLO TURJANSKI⁴ (COORDINADOR)

¹*Universidad de Buenos Aires, Universidad CAECE, AFIP*

²*Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, UBA – CONICET*

³*Departamento de Computación, Facultad de Ciencias Exactas y Naturales, UBA*

⁴*Departamento de Computación, Facultad de Ciencias Exactas y Naturales, UBA – CONICET*

La calidad de datos es un tema que se suele relacionar generalmente con el campo de la industria y la gestión, haciendo referencia principalmente a la confiabilidad y utilidad para la toma de decisiones. Sin embargo, en los últimos tiempos este concepto es cada vez más frecuente en el campo de la investigación, tanto básica como aplicada. La recolección de datos, su posterior curado y cuantificación del impacto de este proceso sobre el análisis subsiguiente, así como potenciales sesgos deberían ser una sección en sí misma de cualquier proyecto de investigación. En esta mesa redonda compartiremos conceptos y experiencias sobre limpieza de datos, documentación, sesgos y confidencialidad, entre otros temas.

Mesa redonda II: Educación estadística

MARIELA MIÑO¹, DANIELA PARADA², LILIANA TAUBER³ (COORDINADORA)

¹*Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

²*Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires – CONICET*

³*Universidad Nacional del Litoral*

La Educación estadística es un área de estudio que, en las últimas décadas, ha ganado terreno tanto en la Educación en general, como en distintos grupos de investigación. A través de la investigación en Educación estadística se han identificado y delimitado diversas componentes (Alfabetización, Razonamiento y Pensamiento estadísticos) que están implícitas en los procesos de enseñanza y de aprendizaje de la Estadística en todos los ámbitos educativos, formales e informales. Asimismo, la investigación en el área, ha permitido mostrar y comprender las complejidades que deberían tenerse en cuenta cuando se diseñan currículos, propuestas didácticas y también, evaluaciones, además de brindar información para reflexionar sobre las formas de valorar las producciones de los estudiantes. Todo este bagaje de conocimiento, ha servido de fundamento para el diseño e implementación sistemática de nuevas propuestas didácticas que permiten introducir los conceptos estadísticos de una manera real, centrada en problemáticas contextualizadas y cercanas a los estudiantes. Así, en esta mesa redonda, se pretende debatir sobre las complejidades más relevantes a la hora de enseñar y aprender Estadística en distintas instancias educativas formales y también, compartir resultados de algunas experiencias.

Ciudad de Buenos Aires, octubre 2024



XXVIII Reunión Científica del Grupo Argentino de Bioestadística





Trabajos Presentados



Análisis Exploratorio y Visualización de datos

ANÁLISIS DE LAS INTERACCIONES HOSPEDADOR-PARÁSITO EN PECES DEL RÍO PARANÁ A TRAVÉS DE REDES BIPARTITAS

MARTÍN ACOSTA ALBARRACÍN^{1,2}, MACARENA DE MARTINO^{1,2}, CLARA VERCELLINI¹, MARTÍN MIGUEL MONTES¹, NATHALIA J. ARREDONDO²

¹ Centro de Estudios Parasitológicos y Vectores (CEPAVE), CONICET, Universidad Nacional de La Plata (CCT-La Plata-CONICET-UNLP), La Plata

² Instituto de Biodiversidad y Biología Experimental y Aplicada (IBBEA, CONICET-UBA) y Departamento de Biodiversidad y Biología Experimental (DBBE, UBA), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

martin.acosta.albarracin@gmail.com

RESUMEN

El uso de redes bipartitas nos permite sintetizar conjuntos de datos complejos como interacciones entre especies, dilucidando patrones estructurales dentro de comunidades ecológicas. El objetivo del trabajo fue analizar y comparar las relaciones entre dos comunidades de peces del Río Paraná y sus endoparásitos a través del uso de redes bipartitas. Mediante una revisión de estudios publicados en Argentina y Brasil, se construyeron dos matrices de presencia-ausencia de interacciones parásito-hospedador separadas en sus respectivas zonas zoogeográficas. Se generaron 100 redes re-muestreadas a través de bootstrap no paramétrico por cada red empírica (función *bootnet* del paquete *bootnet*). Las funciones *networklevel* y *specieslevel* (paquete *bipartite*) se utilizaron para calcular índices que describen la estructura general y de cada especie de la red, respectivamente. Se calculó el estadístico *d* de Cohen a partir del paquete *effsize* para analizar el tamaño del efecto entre los modelos y la red empírica. Por último, los índices generados mediante una prueba de t en R. La red de Argentina (RA) se compuso de 121 parásitos y 65 hospedadores, mientras que la de Brasil (RB) se constituyó de 194 parásitos y 75 hospedadores, presentando 242 y 454 interacciones, respectivamente. El estadístico *d* fue grande ($>0,8$) para todos los índices, excepto para la robustez de hospedadores de RA y la vulnerabilidad de los hospedadores, donde fue pequeña ($<0,3$) en ambas redes. Este último índice fue significativamente ($>0,05$) mayor para la RB. *Genarchella parva* fue el parásito más generalista y central para la estructura de la RA, *Contracaecum* sp. lo fue para la RB y *Pimelodus maculatus* fue el hospedador más vulnerable y estructuralmente más preponderante en ambas redes. La vulnerabilidad de la ictiofauna brasileña permitiría considerarlos como más generalistas. Sin embargo, el sesgo hacia algunos grupos de hospedadores y parásitos en Argentina podría explicar esta diferencia.

Palabras clave: **helminetos, Brasil**

ANÁLISIS ESTADÍSTICO DE LA DINÁMICA DE LA INFECCIÓN POR VIRUS SINCICIAL RESPIRATORIO (VSR) Y ESTIMACIÓN DE SU INCIDENCIA EN EL AMBA

YAMILA ALEN¹, DANIELA PARADA^{1,2}, JULIA DVORKIN^{2,3}

¹ Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires - CONICET

² CONICET (CONICET)

³ Centro INFANT de Medicina Traslacional (CEMIT), Universidad Nacional de San Martín

yami.alen@gmail.com

RESUMEN

A nivel mundial, las infecciones respiratorias agudas bajas (IRAB) son la principal causa de admisiones y hospitalizaciones en unidades de cuidados intensivos pediátricas. En particular, el virus sincicial respiratorio (VSR) es el principal agente etiológico en niños menores de 5 años, incidiendo significativamente en los servicios de salud. El objetivo de este trabajo fue realizar un análisis estadístico descriptivo y estimativo sobre la incidencia de las IRAB por VSR en internaciones pediátricas en el Área Metropolitana de Buenos Aires (AMBA), considerando en particular el impacto de la pandemia COVID-19 en la dinámica de las hospitalizaciones. Con datos proporcionados por el CIMeT, se analizaron 12.767 casos de IRAB pediátricas registrados entre 2018 y 2023 en cuatro hospitales públicos del AMBA. La seroprevalencia de VSR se estimó mediante un diseño observacional, prospectivo y multicéntrico, mientras que para la incidencia se empleó un estudio de cohorte retrospectivo, multicéntrico, en una región pediátrica bien definida. Los resultados cuantifican la carga anual de IRAB severa y crítica por VSR en los periodos estudiados. También se evidencia un retraso del pico de casos en 2021-2022, que puede deberse a las medidas de confinamiento y restricciones de circulación por COVID-19, pero en 2023 retorna a una estacionalidad más temprana, posiblemente por cambios en la dinámica de la infección. A partir de datos poblacionales, se estimó la incidencia según grupo etario y año calendario. Con el aumento de la edad los casos disminuyen independientemente del período analizado. Por otro lado, el análisis mostró el efecto de la pandemia, observándose cambios pre y post pandemia en las tasas de incidencia (IRAB y VSR) según rango etario. Los resultados de este análisis son de gran interés para el desarrollo de políticas públicas de salud, siendo un estudio pionero sobre la incidencia de VSR a nivel poblacional en el AMBA.

Palabras clave: **virus sincicial respiratorio, estadística descriptiva, estimación, incidencia, seroprevalencia, COVID-19.**

Ciudad de Buenos Aires, octubre 2024



DATOS Y ARTE: EN UN ENFOQUE INNOVADOR

BOCA TERESA¹, PUPPO JULIÁN²

¹Departamento de Métodos Cuantitativos y Sistemas de Información, Facultad de Agronomía, Universidad de Buenos Aires

²Artes Multimediales / UNA / Jeyder Multimedia

rtboca@agro.uba.ar

RESUMEN

En un mundo donde las imágenes tienen un impacto creciente, explorar nuevas herramientas que fusionen la ciencia de datos con el arte puede ofrecer un enfoque innovador para la difusión de información. Esta propuesta examina cómo los algoritmos no solo pueden analizar datos, sino también transformarlos en expresiones artísticas que desafían nuestra comprensión tradicional de la estadística. Un aspecto fundamental de este proyecto es el uso del paquete de R *aRtsy*, que permite la creación de arte generativo mediante algoritmos personalizables. Este paquete facilita la conversión de datos estadísticos en visualizaciones artísticas, proporcionando una plataforma flexible y accesible para explorar la creatividad dentro de un marco científico. Al emplear *aRtsy* junto con otras herramientas de software libre, se promueve la democratización del acceso a la tecnología y se fomenta la colaboración interdisciplinaria y la reproducibilidad en distintos contextos educativos y científicos. El enfoque se fundamenta en la visión de *The Creativity Code*, que explora cómo distintos algoritmos pueden transformar los datos en arte generativo. Esta perspectiva destaca el potencial de los algoritmos para innovar en el campo del arte y ofrecer nuevas formas de interpretación de los datos. A través de la generación de un dashboard podemos obtener una herramienta poderosa para la visualización de datos y la estimulación de discusiones, combinando precisión estadística con creatividad artística. La capacidad de compartir el dashboard en línea facilita un acceso inclusivo y una participación activa de diversas audiencias, desde investigadores hasta estudiantes y artistas. Así, el proyecto se alinea con las tendencias actuales en la enseñanza de la estadística y la visualización de datos, inspirando nuevas formas de comunicar y entender los datos y promoviendo una conexión más profunda con las temáticas de interés.

Palabras clave: **visualización, algoritmos, arte generativo, dashborad**

DESCRIPCIÓN Y ANÁLISIS DE LAS POBLACIONES Y COMUNIDADES DE HELMINTOS DE *LAGOSTOMUS MAXIMUS* (RODENTIA, CHINCHILLIDAE)

VICTORIA CANOVA¹, GRACIELA MINARDI¹, AGUSTÍN M. ABBA¹, MARÍA DEL ROSARIO ROBLES¹

¹Centro de Estudios Parasitológicos y de Vectores (CEPAVE, Universidad Nacional de La Plata - CONICET)

victoriac@cepave.edu.ar

RESUMEN

La vizcacha, *Lagostomus maximus*, es un roedor nativo de Sudamérica. El objetivo de este trabajo fue describir y analizar la estructura y composición de las poblaciones y comunidades de helmintos de cuatro poblaciones de vizcachas de diferentes áreas/ambientes: Buenos Aires (SOBA y ECAS), Entre Ríos (ENRI) y Santiago del Estero (SAES). Se consideraron tres niveles de análisis (comunidad componente, infra-comunidad y población componente) describiendo cada nivel mediante variables como prevalencia (P), abundancia (A), abundancia media (AM), intensidad media (IM), riqueza específica (S), diversidad (H'), equitatividad (J') y dominancia (D). Se utilizaron métodos de análisis descriptivos (Análisis de Componentes Principales-ACP) e inferenciales (SIMPER, Hutcheson, Bootstrap, Fisher), ejecutados en R Studio con los paquetes vegan, dabestr, FactoMineR y escolTest. A nivel de comunidad componente, ENRI mostró los mayores valores de A, AM, IM, S y H' , mientras que ECAS presentó el mayor valor de J' y SOBA el mayor valor de D. A nivel de infra-comunidad, el Test de Bootstrap reveló diferencias en A para ENRI, y en S, J' y D para ECAS. SIMPER mostró las mayores disimilitudes entre SAES y los restantes sitios, siendo las especies que más contribuyeron *Viannella* cf. *viscaciae*, *Lagostonema ecasiense* y *Graphidioides* spp. Estas especies, junto con *Heteroxynema* n. sp., *Wellcomia hugoti* y Cestoda, explicaron gran parte de la variabilidad inter- poblacional en el ACP. A nivel de población componente, el Test de Bootstrap mostró diferencias significativas ($p < 0,05$) para las IM y AM de los parásitos más prevalentes entre los sitios considerados. Las estimaciones observadas indican una tendencia que los mayores valores de A, S y H' se asocian con una importante cobertura de vegetación herbácea y arbustiva característica de ENRI, mientras que una menor S y mayor J' se asocia con mayor cobertura de suelo desnudo/urbanizado y vegetación arbórea característico de ECAS.

Palabras clave: *Lagostomus maximus*, helmintos, poblaciones, comunidades, análisis multivariado



ESTADÍSTICA DESCRIPTIVA APLICADA AL MANEJO DE APIARIOS DE LA PROVINCIA DE ENTRE RÍOS

CECILIA FERRUFINO¹, FERNANDA GONZALEZ¹, MARIA JOSE DUS SANTOS^{1,2}

¹*Instituto de Tecnología Agropecuaria (INTA)*

²*CONICET*

ferrufino.cecilia@inta.gob.ar

RESUMEN

En Argentina, Buenos Aires y Entre Ríos son las provincias que albergan mayor número de apiarios destinados a la producción de miel. Entre Ríos presenta una actividad agrícola variada y contribuye a la diversidad floral, fomentando una actividad apícola estacionario o migratoria. Los apicultores implementan prácticas de gestiones cruciales para asegurar la obtención de productos de calidad, aumentar la producción y resguardar la salud de las colmenas. El manejo apícola es evaluado mediante la exploración de las colmenas de apiarios, realizando muestreos e inspecciones para recopilar información en bases de datos que contienen variables de distinta naturaleza y factores tanto fijos como aleatorios que complejizan el análisis de los datos. El objetivo de este trabajo es mostrar la organización y análisis de los datos teniendo en cuenta la hipótesis que se quiere poner a prueba, los diferentes factores y la naturaleza de las variables tomadas durante los muestreos. Se realizaron 4 muestreos en 2 temporadas, colectando datos de un total 326 colmenas, 9 variables respuesta en 2 tipos de manejo apícola. Respecto al análisis estadístico, se llevó a cabo una exhaustiva exploración de datos, realizando los cálculos de valores medios de las variables cuantitativas continuas y porcentajes de las categorías de las variables cualitativas para cada uno de los factores. Además se utilizó la prueba de chi-cuadrado para comparar los diferentes tipos de manejo, temporada y muestreo para cada una de las variables categóricas. Como resultado se observaron diferencias significativas según el tipo de manejo, temporada y muestreo para las variables nutrición, categoría de colmena y nivel de infección de un microsporidio. Respecto a los análisis estadísticos este trabajo desarrolla una metodología para ordenar, analizar e interpretar datos de una manera sencilla evaluando todas las variables de manera integral para contribuir al éxito y la sostenibilidad de la actividad apícola.

Palabras clave: ***estadística descriptiva, actividad apícola, Entre Ríos***

HIPERGRAFOS: UN NUEVO ENFOQUE METODOLÓGICO PARA DETERMINAR ESPECIES CLAVE EN UN ECOSISTEMA

HUAYLLA CLAUDIA¹, KUPERMAN MARCELO², GARIBALDI LUCAS¹

¹CONICET. Instituto de Investigaciones en Recursos Naturales, Agroecología y Desarrollo Rural. Universidad Nacional de Río Negro. San Carlos de Bariloche, Río Negro.

²Centro Atómico Bariloche (CNEA) and CONICET. Instituto Balseiro, Universidad Nacional de Cuyo, San Carlos de Bariloche, Río Negro.

cahuaylla@unrn.edu.ar

RESUMEN

Actualmente la teoría de grafos es ampliamente usada en diferentes disciplinas. Muchos sistemas pueden ser representados mediante un grafo. En los últimos años se han realizado avances en el área de la ecología, ya que los grafos pueden describir la estructura de una amplia variedad de sistemas complejos, especificando qué pares de entidades del sistema están conectados. Cuando las interacciones involucran a más de dos entidades al mismo tiempo surge la necesidad de usar otras herramientas como los hipergrafos. Los hipergrafos son generalizaciones de grafos cuyas aristas contienen más de un nodo y pueden representar interacciones simultáneas entre varias unidades. En este trabajo se usó un conjunto de datos para modelar un ecosistema en Patagonia Norte. Los tipos de nodos involucrados son especies de polinizadores, especies de plantas y gremios de herbívoros. Estos fueron representados por un hipergrafo donde cada hiperenlace está compuesto por especies de plantas y gremios de herbívoros, los cuales interactúan con una especie de planta. Los hipergrafos pueden ser caracterizados usando diversos índices, en este trabajo se usó la entropía para calcular la efectividad de cada hiperenlace dentro del sistema total, donde cada hiperenlace representa una especie de planta. Para el cálculo de este índice se creó una función en RStudio, el mismo puede ser usado para cualquier hipergrafo. Los resultados obtenidos arrojaron que las especies de plantas más relevantes en el sistema son *R. magellanica* y *E. coccineum*. Sus valores de entropía fueron de 2.89 y 2.94 respectivamente, siendo estos los valores más altos dentro del conjunto de plantas. Curiosamente, estos resultados se corresponden con un análisis previo realizado con este conjunto de datos, pero en el cual se usó otra metodología, llegando a los mismos resultados. Este análisis es eficiente para identificar nodos que tienen más influencia dentro del sistema de estudio.

Palabras clave: *ecosistema, entropía, interacciones ecológicas.*



Análisis Multivariado



COMPARACIÓN DE ALGORITMOS NIPALS Y SVD PARA REGRESIÓN POR MÍNIMOS CUADRADOS PARCIALES EN GWAS

EUGENIA BORTOLOTTO¹, GERARDO CERVIGNI², CECILIA BRUNO³

¹Doctorado en Estadística, Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario, Rosario

²Centro de Estudios Fotosintéticos y Bioquímicos (CEFOBI-CONICET), Rosario

³Cátedra de Estadística y Biometría. Facultad de Ciencias Agropecuarias. Universidad Nacional de Córdoba. Grupo vinculado de Estadística y Biometría Unidad de Fitopatología y Modelización Agrícola (UFyMA-INTA-CONICET), Córdoba

bortolotto.eugenia@gmail.com

RESUMEN

El método de mínimos cuadrados parciales (PLS) es una regresión que utiliza variables latentes para modelizar la estructura de dos espacios factoriales y maximizar la relación existente entre ambos. Para la estimación de variables latentes usa dos algoritmos: 1. NIPALS (Non-linear Iterative Partial Least Squares) recomendado para matrices de alta dimensión y con valores faltantes, ya que puede calcular las componentes latentes una por una. Es eficiente en términos de memoria computacional y permite un análisis iterativo; 2. SVD (Descomposición del Valor Singular) que descompone una matriz en tres: $X = U\Sigma V'$ y proporciona una solución exacta asegurando que se capturen todas las variaciones importantes en los datos. Este método es robusto y preciso matemáticamente pero computacionalmente más intensivo. El objetivo de este trabajo es evaluar y comparar el rendimiento de los algoritmos NIPALS y SVD en la solución de regresiones-PLS. Se aplicaron ambas descomposiciones sobre dos conjuntos de datos caracterizados fenotípicamente y genómicamente con marcadores SNP. Una base de datos simulada con 80K SNPs y una variable fenotípica medida en dos ambientes y una base de datos real de duraznero con 619 SNPs y tres caracteres fenotípicos medidos en dos campañas (ambientes). Cada base de datos se dividió en un conjunto de entrenamiento y uno de prueba (70% y 30%, respectivamente). Sobre el conjunto de entrenamiento se realizó un PLS para modelar la relación entre las variables moleculares (predictoras) y fenotípica (respuesta). El conjunto de prueba se utilizó para evaluar el comportamiento y generalización del modelo-PLS. La capacidad predictiva se evaluó con la raíz del error cuadrático medio (RECM). Los resultados fueron visualizados en un triplot. El algoritmo NIPALS arrojó mejores resultados de RECM respecto a SVD en ambas bases de datos ($RECM_{NIPALS}^{sim} = 9$, $RECM_{SVD}^{sim} = 20$, $RECM_{NIPALS}^{real} = 33$, $RECM_{SVD}^{real} = 2125$) y además identificó mayor cantidad de marcadores significativos que SVD.

Palabras clave: **Regresión-PLS, SNPs, fenotipos, genotipos, mapeo asociativo, métodos multivariados**



ENTRENAMIENTO COGNITIVO DE FUNCIONES EJECUTIVAS EN NIÑOS ARGENTINOS: APLICACIÓN DE MÉTODOS DE CLUSTERING PARA EL ANÁLISIS DE PERFILES DE RENDIMIENTO

MARTINA BOSCOLO¹, GABRIEL O. PAZ¹, DIEGO E. SHALOM^{2, 3}, MARIA JULIA HERMIDA⁴, ANDREA P. GOLDIN¹

¹ Laboratorio de Neurociencia. Centro de Inteligencia Artificial y Neurociencia (CIAN). Escuela de Negocios. Universidad Torcuato Di Tella. CONICET

² Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Buenos Aires

³ Instituto de Física Interdisciplinaria y Aplicada (INFINA), CONICET—Universidad de Buenos Aires

⁴ Universidad Nacional de Hurlingham - CONICET (UNAHUR-CONICET), Villa Tesei, Buenos Aires

martiboscolo@gmail.com

RESUMEN

Las funciones ejecutivas (FE) son funciones cognitivas que permiten controlar acciones y pensamientos y adaptarse a entornos cambiantes. Son importantes para el éxito educativo y de la vida, y pueden mejorarse con entrenamiento cognitivo. Desde hace más de quince años se implementa Mate Marote, un *software* lúdico de acceso gratuito para entrenamiento y evaluación de FE en niños de 4 a 8 años. Las intervenciones duran de 1 a 4 meses y ocurren dentro de las escuelas, con resultados exitosos. En este estudio se presenta el análisis de una intervención realizada con niños argentinos de 6 años. Se evaluó su rendimiento en FE antes y después de un entrenamiento de alrededor de 27 sesiones de 15 minutos de juego cada una distribuidas a lo largo de 3 meses, y se lo comparó con el de un grupo control activo. El objetivo del presente trabajo fue realizar análisis de *clustering* de estos datos para determinar cómo se agrupan los participantes según su rendimiento en las pruebas y cómo se relaciona el rendimiento cognitivo entre las distintas FE, con el fin de contribuir al entendimiento del entrenamiento cognitivo personalizado. Se utilizó el método *k-means* y el método de la silueta para la validación de la cantidad óptima de clusters sobre los datos del grupo experimental. Además, se caracterizaron y compararon los grupos en cuanto a variables sociodemográficas y académicas. Se discute la interpretación de los resultados en relación con las implicancias para el análisis de perfiles de rendimiento en el entrenamiento cognitivo; así como la relevancia y ventajas de aplicación de métodos de *clustering* para el análisis de datos en Ciencias Cognitivas y del Comportamiento.

Palabras clave: **entrenamiento cognitivo, entrenamiento personalizado, clustering, k-means.**

EVALUACIÓN DE INDICADORES METABÓLICOS EN CORDEROS TEXEL EMPLEANDO MODELOS LINEALES MIXTOS

CENTURIÓN N.^{1,2}, FERREYRA S.^{1,2}, JULIANO N.⁴, GALARZA R.⁵, IORIO J.⁴, ROVEGNO S.³, MICCOLI F.^{3,4},
ABBIATI N.³

¹CITNOBA (CONICET- UNNOBA-UNSAa)

²Universidad Nacional de San Antonio de Areco (UNSAa)

³Facultad de Ciencias Agrarias, Universidad Nacional de Lomas de Zamora

⁴Facultad de Agronomía, Universidad de Buenos Aires

⁵INTA Cuenca del Salado

ncenturion@agrarias.unlz.edu.ar

RESUMEN

En ovinos, la suplementación lipídica es una estrategia para incrementar los ácidos grasos poliinsaturados (PUFA) y omega 3 en la carne. Sin embargo, podría haber cambios en el metabolismo lipídico a nivel hepático. El objetivo fue evaluar indicadores de metabolismo lipídico en corderos cruce Texel suplementados con semilla de lino. Se seleccionaron 24 corderos cruce Texel de 45 días (12 hembras y 12 machos) armando 12 bloques según peso y sexo. Los animales se encontraban en crianza 100% pastoril hasta el día 1 del ensayo, a partir del cual se definieron dos tratamientos dietarios durante 72 días: 1) Concentrado (CON) y Concentrado con semilla de Lino al 6% (LIN). Se alojaron en corrales individuales y se tomaron muestras de sangre al día 1 y 72. Para el análisis estadístico, se consideraron dos grupos de variables: G1= enzimas hepáticas (GOT, FAS, GGT, CPK y LDH) y G2= lípidos (Triglicéridos, Colesterol, HDL, LDL, ácidos grasos no esterificados (NEFA) y beta hidroxibutirato (BHB)). Se realizó un análisis exploratorio de Componentes Principales (ACP) para cada grupo. Luego, se emplearon modelos lineales mixtos con bloques aleatorios y un arreglo factorial de tratamientos y días, eligiéndose la correlación y la heterocedasticidad en función del criterio de Akaike. Se trabajó con el software Navure con $\alpha=0,05$. No se detectaron interacciones significativas. Para G1 se detectaron diferencias sólo entre días de muestreo, sin verse afectado el metabolismo hepático por la inclusión de semilla de lino. Para G2 en Triglicéridos se encontraron diferencias tanto en días como en tratamiento, siendo mayor el promedio en LIN; para BHB no hubo diferencias y en el resto de las variables diferencias entre días. Los resultados fueron similares a lo observado en los Biplots de ACP.

Palabras clave: ***corderos, suplementación lipídica, metabolismo hepático, calidad de carne.***

PROTOCOLO DE ZONIFICACIÓN Y CARACTERIZACIÓN EDAFOCLIMÁTICA. APLICACIÓN EN UN VALLE VITIVINÍCOLA

CORDOBA M.^{1,2}, PACCIORETTI P.^{1,2,3}, BALZARINI M.^{1,2}

¹ *Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba*

² *Grupo de Estadística Vinculado. Unidad de Fitopatología y Modelización Agrícola. (UFyMA-INTA-CONICET). Córdoba*

³ *Departamento de Agronomía y Horticultura, Universidad de Nebraska.*

mariano.cordoba@unc.edu.ar

RESUMEN

Comprender la variación espacial dentro de una región vitivinícola es fundamental para describir las características específicas del ambiente que influyen en la calidad del vino, lo que implica la delimitación de zonas homogéneas en términos de clima, geomorfología y suelos. En la zonificación edafoclimática, es crucial considerar la co-variación espacial de múltiples factores que influyen en la variabilidad y calidad de los viñedos. Este trabajo utiliza un enfoque metodológico que integra diversas herramientas estadísticas avanzadas para identificar zonas edafoclimáticas homogéneas. Primero, se delimitan subzonas bioclimáticas utilizando el método KM-sPC, que combina el Análisis de Componentes Principales Espaciales (sPCA) con el análisis de conglomerados fuzzy K-means. Luego, utilizando el mismo método, se realiza una zonificación del suelo dentro de cada subzona climática, agrupando sitios con alta similitud en variables edáficas y geomorfológicas. La caracterización de las zonas homogéneas se efectuó mediante gráficos de estrellas, evaluando la importancia de las variables en la clasificación de cada zona mediante el algoritmo Random Forest. Las zonas delimitadas se evaluaron comparando, para cada variable, las medias entre zonas utilizando un método basado en permutaciones que considera la correlación espacial en los datos. El proceso analítico se ilustra con datos del valle vitivinícola del río Mendoza. Los resultados muestran que la delimitación de zonas homogéneas proporciona una base sólida para establecer la tipicidad de los ambientes vitivinícolas en las regiones productoras de vino. La metodología empleada asegura la coherencia espacial de las zonas delimitadas y se basa en las variables más influyentes que explican la co-variabilidad edafoclimática.

Palabras clave: **Análisis de clúster espacial, Zonas edafoclimáticas, Importancia de variables.**

ANÁLISIS DE REDUNDANCIA CON PARTICIÓN DE LA VARIANZA: UNA HERRAMIENTA ÚTIL PARA ESTUDIAR LA INFLUENCIA DE MÁS DE UN CONJUNTO DE VARIABLES AMBIENTALES EN LA VARIABILIDAD DE UNA COMUNIDAD DE AVES

CORRAL, MARÍA GABRIELA¹ Y FERNÁNDEZ, GUSTAVO JAVIER²

¹*Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

²*INIBIOMA, Centro Regional San Martín de los Andes, Universidad Nacional del Comahue*

gabicorral@ege.fcen.uba.ar

RESUMEN

Los pastizales son uno de los ecosistemas más transformados del mundo por el desarrollo productivo. La identificación de los atributos del hábitat que influyen en la abundancia de las aves puede guiar estrategias de manejo. La composición y estructura de la vegetación puede afectar directamente la supervivencia y reproducción de las aves. El objetivo de este trabajo fue estudiar los factores ambientales asociados a los cambios introducidos por la ganadería que influyen la presencia y abundancia de aves del pastizal salobre del sur de la Bahía de Samborombón. A partir de conteos de aves se obtuvo la abundancia relativa de cada especie observada en 38 transectas. Mediante muestreos de vegetación se obtuvieron dos grupos de variables: 1) porcentaje de cobertura vegetal por especie y 2) descriptores de altura. Para estudiar la influencia de estas variables sobre la comunidad de aves se realizó un Análisis de Redundancia (RDA) y para determinar la importancia de cada grupo se realizó una Partición de la Varianza. Este análisis multivariado permite calcular qué porcentaje de la variación en la matriz de aves es explicada por cada grupo de variables de vegetación y cuánto queda sin explicar. La variación pura explicada por el grupo altura representó un 13.4%, la variación compartida por los descriptores de altura y el tipo de cobertura vegetal representó un 16.4% y la variación pura debida al tipo de cobertura vegetal representó un 2.1%. La variación debida al efecto conjunto de ambos grupos fue significativa como también lo fue la variación pura debida a los descriptores de altura. Sin embargo, la variación pura debida a las variables de cobertura no fue significativa. De esta manera, se pudo determinar que las variables más influyentes en la comunidad de aves estuvieron asociadas a la estructura de la vegetación y no a la composición florística.

Palabras clave: ***RDA, partición de la varianza, aves de pastizal, ganadería, estructura de la vegetación***

PATRONES DE EXPRESIÓN EN PROTEÍNAS DE LA FAMILIA HSP20 EN ESTRÉS ABIÓTICO REVELADOS POR ANÁLISIS FACTORIAL MÚLTIPLE JERÁRQUICO

COSTA TÁRTARA, S.^{1,2}; ARCE, D.^{2,3}; TOLOSA, G.¹; PRATTA, G.^{2,4}

¹*Departamento de Ciencias Básicas, Universidad Nacional de Luján.*

² *CONICET (CONICET)*

³ *Secretaría de Ciencia y Tecnología, Facultad Regional San Nicolás, Universidad Tecnológica Nacional, San Nicolás, Buenos Aires*

⁴ *Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Zavalla, Santa Fe*

scosta@unlu.edu.ar

RESUMEN

Los datos ómicos representan los sistemas biológicos que conforman el funcionamiento celular y, por ende, presentan diferentes formatos en función de su naturaleza. Se cuenta con un conjunto de datos construido desde bases públicas e información reportada en experimentos independientes. La matriz contiene 52 genes codificantes de proteínas de choque térmico de bajo peso molecular (Hsp20) identificados en el genoma de quinoa, y en transcriptomas de diferentes tejidos, expuestos a condiciones de estrés abiótico: salinidad (S), alta temperatura (H) y déficit hídrico (D). Dado que cada estrés induce la expresión de diferentes Hsp20, se optó por aplicar Análisis Factorial Múltiple Jerárquico (AFMJ) (implementado en FactoMineR) para identificar patrones y agrupamientos de los genes en relación a la ocurrencia de estrés. Se recuperaron datos de expresión diferencial (ED) normalizados y se consideró 0 (cero) para los casos que no presentaron ED. La jerarquía se determinó por el efecto de los estreses sobre la ED de los genes. Complementariamente, se consideraron tres variables cualitativas que clasifican las Hsp20 por su peso molecular, la localización subcelular del producto proteico y la ocurrencia en S, H y/o D y sus combinaciones. Los dos primeros ejes del AFMJ acumularon el 23,2 % de la varianza; las variables con mayor contribución fueron la ED en S y H para el primer eje, y de D para el segundo. Se observan dos grupos de genes en DS y S, que traducen proteínas en cloroplasto y mitocondria, mientras que las Hsp20 cuyo producto se localiza en el núcleo se presentan en el resto de las condiciones. La aplicación de AFMJ representa un enfoque multivariado exploratorio sobre Hsp20 de quinoa. Es necesario validar su transferibilidad a conjuntos que contengan mayor número de genes compartidos entre condiciones, a fin de evaluar la capacidad del método de capturar mayor variabilidad.

Palabras clave: ***estrés abiótico - genes - expresión - quinoa - HMFA***

ANÁLISIS DE SERIES TEMPORALES PARA ESTUDIAR LA RELACIÓN ENTRE LA SEQUÍA Y LA VEGETACIÓN EN LOS BOSQUES SUBTROPICALES DE MISIONES

DE LELLIS RM¹, DIAZ VILLA MVE¹², PARONETTO JS¹², CAGNONE MF¹², CRISTIANO PM¹²

¹ *Laboratorio de Ecología Funcional, Departamento de Ecología Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

² *Instituto de Ecología, Genética y Evolución (IEGEB), Universidad de Buenos Aires-CONICET*

rociodelellis@gmail.com

RESUMEN

El análisis estadístico de las series temporales puede utilizarse para comprender la dinámica de la interacción entre un conjunto de variables. El objetivo principal de este trabajo fue evaluar el efecto de la sequía en la vegetación de los bosques subtropicales de Misiones, utilizando series temporales del índice de vegetación EVI (*Enhanced Vegetation Index*) y el índice de sequía PDSI (*Palmer Drought Severity Index*), desde el año 2000 al 2023. Para analizar las variaciones conjuntas entre los índices, se ajustó un modelo de autorregresión vectorial (VAR) multivariado, al cual luego se le aplicó un Test de Causalidad de Granger para evaluar la capacidad predictiva entre los índices. Se verificó la estacionariedad de las series temporales utilizando tanto el test de Phillips-Perron (PP) como el test de Kwiatkowski, Phillips, Schmidt y Shin (KPSS). Los resultados del test de causalidad de Granger indicaron que los valores pasados y presentes del PDSI no ayudan a mejorar la predicción del EVI, en comparación con un modelo que utiliza únicamente los datos del EVI. Dado que la prueba no mostró relación entre el EVI y el PDSI, en este contexto se concluye que no hay efectos significativos de la sequía en la actividad fotosintética de la vegetación en el período estudiado. Aún así, en un ecosistema sin limitación de agua como recurso y una densa vegetación con múltiples estratos, las variaciones temporales en el EVI pueden ser muy sutiles y difíciles de detectar, por lo que comprender esta relación puede ser complejo.

Palabras clave: **modelo VAR, causalidad, serie temporal multivariada, autocorrelación**

CALIDAD DE LECHE POR GRUPOS ETÁREOS EN BOVINOS CRIOLLOS ARGENTINOS

ERNETA, L.¹, TOPAYAN, MV.^{1,2,3}, GÉNERO, E.^{2,3}, ABBIATI, N.^{2,3}

¹Facultad de Ciencias Agrarias, Universidad Nacional de Lomas de Zamora

²Instituto de Investigación sobre Producción Agropecuaria, Ambiente y Salud (IIPAAS – CIC)

³Comisión Nacional de Investigaciones Científicas y Técnicas (CONICET)

luciana.erneta@hotmail.com

RESUMEN

La calidad en leche bovina depende de atributos fisicoquímicos, higiénicos y organolépticos que determinan su capacidad de satisfacer necesidades, en la alimentación del ternero o en su uso como producto alimenticio para consumo humano. La leche está formada por más de 100 sustancias que se encuentran en solución, suspensión o emulsión; entre ellas lactosa, proteínas, grasa y vitaminas solubles. El objetivo del trabajo fue describir la calidad láctea en un rodeo de bovinos criollos argentinos y su variación con la edad de los animales, empleando un analizador ultrasónico. Se obtuvieron muestras aleatorias de leche de un rodeo experimental de hembras bovinas de raza criolla con cría al pie (n=19), clasificadas en tres grupos etarios (GE): jóvenes (≤ 10 años), medias (11-12 años) y adultas (≥ 13 años). Se empleó el analizador ultrasónico Lactoscan MCC. Se midieron 9 atributos cuyo estudio comprendió análisis de correlaciones (AC) y componentes principales (ACP) con variables estandarizadas. Para aquellas sin multicolinealidad, se efectuó análisis de varianza (ANVA) para comparar las medias por GE. Por ACP, Se descartaron 5 atributos que resultaron combinaciones lineales, permaneciendo: porcentaje de grasa, densidad, temperatura de la muestra y punto de congelación, empleados en ACP. Los dos primeros componentes explican el 83% de la variabilidad, sugiriendo que la mayoría de la información relevante sobre la calidad láctea podría ser capturada a través de estas dos dimensiones. El *biplot* sugiere una dispersión más notoria en edades avanzadas, sin observarse diferencias entre medias. Los ANVA corroboraron al *biplot*. Esto indicaría que, para la población bajo estudio, la variabilidad de los atributos de calidad dependería de la edad de los animales, pero no su promedio. Los hallazgos podrían ser utilizados para futuras investigaciones que busquen optimizar prácticas de manejo y alimentación en rodeos de cría. Los resultados fueron obtenidos empleando los softwares InfoStat y R.

Palabras clave: **cría bovina, lactoscan, análisis de varianza, densidad, porcentaje de grasa en leche**

USO DE COVARIABLES CLIMÁTICAS Y EDÁFICAS PARA LA DEFINICIÓN DE “AMBIENTE” EN ESTUDIOS DE ESTABILIDAD DEL RENDIMIENTO DE MAÍZ EN ARGENTINA

FEDERICO IZZO¹, CRISTINA CUESTA²

¹*Breeding, Bayer*

²*Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario*

federico.izzo@bayer.com

RESUMEN

La presencia de interacciones genotipo por ambiente (IGA) es uno de los principales obstáculos para la selección de los mejores genotipos en un programa de mejoramiento genético. Dichas interacciones dan lugar a cambios de ranking para el carácter analizado según el ambiente explorado (interacciones cualitativas). Numerosos autores han propuesto índices para cuantificar la IGA, dando lugar a estudios llamados “análisis de estabilidad”. Esos estudios se agrupan según la metodología estadística utilizada para derivar el índice, siendo las más populares las basadas en métodos de regresión lineal, análisis multivariados o modelos mixtos. El carácter más estudiado para estabilidad en maíz es el rendimiento. En gran parte de los trabajos realizados sobre la estabilidad de este carácter, el factor ambiente es determinado como el promedio del rendimiento de todos los genotipos en cada ensayo comparativo de rendimiento (ECR) realizado. Esto tiene consecuencias en el análisis estadístico (falta de cumplimiento de supuestos en análisis de regresión, por ejemplo) y en la capacidad de dar sentido conceptual a la interacción, ya que un mismo promedio de rendimiento puede ser el resultado de combinaciones de niveles de factores climáticos y edáficos muy diferentes. El objetivo de este trabajo es definir “ambiente” de una forma que provea bases conceptuales para el estudio de la IGA. Para esto, se utilizan datos climáticos y edáficos de 36 localidades repetidas en 3 campañas agrícolas de la región templada de Argentina. Como técnica estadística, se propone el uso de análisis de clusters. Los resultados indican que la metodología podría ser aplicada con éxito para el análisis de la IGE. Los clusters resultantes representan ambientes homogéneos (por ejemplo, la primera agrupación fue debido a fechas de siembra, dividiendo ambientes principalmente por temperatura y evapotranspiración), brindando la oportunidad de analizar la IGA con una interpretación simplificada y con sentido biológico.

Palabras clave: ***Interacción genotipo por ambiente; análisis de cluster; estabilidad***



DETECCIÓN DE ASOCIACIONES GENÉTICAS EN ACCESIONES DE QUÍNOA DEL NOROESTE ARGENTINO MEDIANTE MAPEO ASOCIATIVO CON VARIABLE DEPENDIENTE CUALITATIVA

LAC PRUGENT VERONICA ¹, COSTA TARTARA SABRINA ², CURTI RAMIRO ³, BRAMARDI SERGIO ^{1,4}

¹ *Departamento de Estadística, Universidad Nacional del Comahue, Neuquén*

² *Departamento de Tecnología, Universidad Nacional de Luján*

³ *Laboratorio de Investigaciones Botánicas, Universidad Nacional de Salta, Salta*

⁴ *Centro de Investigaciones en Toxicología Ambiental y Agrobiotecnología del Comahue (UNCo-CONICET)
lacprugentvero@gmail.com*

RESUMEN

En las últimas décadas, el cultivo de la quínoa ha despertado un creciente interés debido a su elevado valor nutritivo. La variabilidad genética de la quínoa permite el desarrollo de variedades adaptables a diversas condiciones agroecológicas, consolidándola como una opción viable frente a los desafíos alimentarios globales. En estos procesos de mejora es de gran importancia establecer relaciones entre caracteres fenotípicos y marcadores moleculares, lo que se conoce como mapeo asociativo, ampliamente utilizado cuando la variable dependiente es cuantitativa. El objetivo del presente trabajo es identificar modelos de mapeo asociativo entre caracteres fenotípicos cualitativos y marcadores moleculares en quínoa. En este estudio, se analizaron 25 accesiones provenientes de la región del Noroeste Argentino, evaluando seis variables fenotípicas dicotómicas (tipo y forma de la panoja, presencia de ramas, axilas, pigmentos en cotiledones y estrías) y 313 variables moleculares correspondientes a 22 loci microsatélites (SSR). Previo al desarrollo de los modelos, se evaluó el desequilibrio de ligamiento (LD) de los marcadores moleculares utilizando la matriz de correlaciones al cuadrado. Se observó que el 1,22% de las correlaciones superaron el umbral de 0,8 y se rechazó la hipótesis de no correlación mediante el test de esfericidad de Bartlett. Para cada variable cualitativa binaria, se ajustaron modelos de regresión logística sin estructura con cada uno de los marcadores moleculares resultando 313 modelos. La significación de los marcadores se evaluó aplicando la corrección de Benjamini-Yekutieli (2001) para reducir los descubrimientos falsos positivos. Sólo un marcador mostró asociación significativa con la forma de la panoja (Amarantiforme o Glomerular). Un test de independencia Chi-cuadrado confirmó esta relación. El análisis del modelo de regresión logística indicó que la presencia del marcador está asociado al tipo de panoja amarantiforme. De esta manera a partir de la metodología propuesta se logró identificar asociación entre un carácter cualitativo y un marcador molecular.

Palabras clave: **Modelos de Regresión Logística, Corrección de Benjamini-Yekutieli, Desequilibrio de Ligamiento**



CLASIFICACIÓN DE LA CALIDAD DEL HÁBITAT EN CURSOS DE AGUA DE LA PROVINCIA DEL NEUQUÉN A PARTIR DEL ÍNDICE IVH

LAFFITTE LORENA¹⁴, ABELLI BONARDI MARISOL², FERNÁNDEZ JOSÉ CARLOS¹,
GARCÍA TOURN ANDREA², LAURENZANO BETINA², LILLO MANUEL¹, MUÑIZ
SAAVEDRA JULIETA⁴, RIVERA DANIEL¹, TORRES MARIANA¹, TRONCOSO
LEANDRO², UGOLINI FLORENCIA² Y LAVALLE ANDREA³

¹ *Secretaría de Ambiente de la Provincia del Neuquén*

² *Subsecretaría de Recursos Hídricos de la Provincia del Neuquén*

³, *Departamento de Estadística, Facultad de Economía, Universidad Nacional del Comahue,
Neuquén*

⁴ *CRUSMA, Universidad Nacional del Comahue, Neuquén*

laffittesmandes@gmail.com

RESUMEN

El trabajo tuvo como objetivo establecer categorías de calidad del hábitat, a partir de la aplicación del Índice de Valoración del Hábitat (IVH) en cursos de agua de la provincia del Neuquén. Se utilizó información de 12 años de muestreos realizados en 93 sitios distribuidos en las cuencas de los ríos Limay, Neuquén, Hua Hum y Colorado. Se incluyeron variables fisicoquímicas como conductividad eléctrica, oxígeno disuelto, nitrógeno total, fósforo total y alcalinidad, además de las 10 variables relacionadas con la caracterización del hábitat y el índice biótico BMPS. Para determinar los rangos que definen las categorías de calidad del IVH, se aplicó un Análisis de Componentes Principales (ACP) utilizando 16 variables descriptivas. La primera componente principal explicó el 39,08% de la variabilidad total y mostró una correlación de 0,96 con el IVH, lo que permitió clasificar los sitios en cuatro grupos. Se establecieron los siguientes puntos de corte para las categorías de calidad del hábitat: 0 a 70 (pobre), 71 a 120 (marginal), 121 a 139 (subóptimo) y 140 a 200 (óptimo). La clasificación fue validada mediante un Análisis Discriminante, obteniéndose una tasa de error del 4%, lo que respalda la robustez de los resultados. La clasificación de la calidad del hábitat en diferentes categorías proporciona una herramienta que permite identificar su condición general y realizar un seguimiento más claro de la salud de los ambientes acuáticos a lo largo del tiempo.

Palabras clave: ***análisis multivariado, valoración del hábitat, índice, macroinvertebrado***



DESARROLLO Y EVALUACIÓN DE UNA METODOLOGÍA DUAL PARA DATOS CATEGÓRICOS BASADA EN EL ANÁLISIS FACTORIAL MÚLTIPLE

ANDREA LAVALLE¹, LISANDRO CURIA²

¹*Departamento de Estadística, Universidad Nacional del Comahue*

²*Departamento de Matemática, Universidad Nacional del Comahue*

lavalleandrealina@gmail.com

RESUMEN

En los estudios de caracterización con variables son categóricas, puede ocurrir que los individuos están previamente clasificados en grupos. En este contexto es de interés comparar si las asociaciones existentes entre las variables son similares en los diferentes grupos. Las estructuras que poseen una partición de individuos, pueden estudiarse a través de técnicas duales de análisis multivariado a tres vías. En el caso de variables cuantitativas, se puede aplicar el Análisis Factorial Múltiple Dual (AFMD). En este trabajo se propone una metodología para el caso de variables categóricas basada en el análisis de datos mixtos desarrollado por Bécue Bertaut y Pagès (2007). En el mencionado desarrollo los autores tratan a los grupos de variables categóricas de la siguiente manera: se obtiene la tabla disyuntiva completa (TDC); cada columna se pondera por el inverso de la proporción de individuos que poseen la categoría; la tabla ponderada se centra; por último se realiza un Análisis de Componentes Principales no normado. La metodología propuesta en este trabajo consiste en realizar el mismo tratamiento descripto a cada tabla correspondiente a un grupo de individuos, pero realizar el ACP no normado sobre la concatenación de todas las tablas, siguiendo la idea del AFMD. De esta forma, se respeta la estructura interna de cada tabla y el centrado permite la concatenación. Al realizar el ACP con todas las tablas, se obtiene la estructura consenso, que refleja el comportamiento de las variables en el grupo completo de individuos, y provee las coordenadas de los individuos. Las modalidades parciales se obtienen proyectando las subtablas en el consenso. Se realiza una aplicación de la técnica y se compara con otras propuestas para estructuras duales categóricas, desarrolladas por Abascal et al. (2012) y Fernández Aguirre et al. (2013).

Palabras clave: ***caracterización, conjuntos de individuos, asociación.***

APORTES DEL ANÁLISIS FACTORIAL MÚLTIPLE DUAL MIXTO EN EL ESTUDIO DE ESTRUCTURAS DE DATOS MULTIVÍA

DEL MEDICO A. P.¹, VITELLESCHI M. S.², PRATTA G.¹, LAVALLE A.³

¹*Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR)*

²*Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística (IITAE),
CIUNR-Universidad Nacional de Rosario*

³*Departamento de Estadística, Universidad Nacional del Comahue*

lavalleandrealina@gmail.com

RESUMEN

El Análisis Factorial Múltiple (AFM) en el análisis de datos mixtos trata a los individuos como primera vía y a los grupos de variables de igual naturaleza como la segunda y tercera vía. En una estructura dual, con partición de individuos, se añade una cuarta vía. Se desarrolló el Análisis Factorial Múltiple Dual Mixto (AFMDMix), combinando los fundamentos del AFM dual y mixto, estandarizando y concatenando variables cuantitativas, y transformando y concatenando tablas disyuntivas completas. Éstas se yuxtaponen en una matriz sobre la cual se realiza un Análisis de Componentes Principales (ACP) no normado, obteniendo coordenadas de individuos, variables parciales y consenso. Se comparó el AFMDMix con el AFM mediante datos correspondientes a 13 familias provenientes de la autofecundación de las generaciones F4 y F5 de tomate (*Solanum lycopersicum*), evaluadas en 12 atributos cuantitativos de calidad y 3 marcadores moleculares InDel, para los cuales existen la forma C (cultivado, cv. Caimanta) y P (silvestre, LA0722 de *S. pimpinellifolium*). En F4, se observan asociaciones fuertes entre variables como altura, peso, diámetro y reflectancia. Las familias con perfiles de InDel de forma C se asocian con valores bajos en altura, peso y diámetro, pero altos en acidez e índice chroma. En F5, se presenta un cambio en las relaciones, destacando las asociaciones entre reflectancia y vida postcosecha. Las familias segregantes para los perfiles de InDel (es decir, que presentan tanto la forma C como la P) se asocian con valores más altos de acidez pero bajos de índice de madurez. A diferencia del AFM clásico, el AFMDMix captura estas diferencias, permitiendo una comprensión más profunda de las relaciones fenotípicas y genotípicas entre generaciones, esencial para el mejoramiento genético. En conclusión, el AFMDMix resulta una herramienta útil para el análisis de datos complejos en estudios genéticos, ofreciendo ventajas significativas sobre los métodos tradicionales.

Palabras clave: **Estructuras duales, Tomate (*Solanum lycopersicum*), caracterización fenotípica, análisis factorial múltiple**

EVALUACIÓN DE BIOESTIMULANTE A BASE DE ALGAS EN PLANTINES DE ALHELÍ (*MATTHIOLA INCANA* (L.) R.BR.)

MERGEL, CA¹, MILICIA, V², ABBIATI, NN²

¹Facultad de Ciencias Agrarias/ Universidad Nacional de Lomas de Zamora

²Instituto de Investigación sobre la Producción Agropecuaria, Ambiente y Salud (IIPAAS). Facultad de Ciencias Agrarias, Universidad Nacional de Lomas de Zamora

cmergel@agrarias.unlz.edu.ar

RESUMEN

Actualmente, existe una creciente tendencia mundial a utilizar bioestimulantes orgánicos en la agricultura debido a los beneficios que producen en las plantas, favoreciendo el crecimiento y promoviendo el desarrollo fisiológico. Se realizó un ensayo con un bioestimulante a base de algas con tres alternativas de aplicación: radicular, foliar, mixto, y un control, en la especie floral alhelí con el objetivo de evaluar sus efectos. El diseño experimental fue completamente aleatorizado con tres repeticiones y 9 submuestras (plántulas). Se sembraron 36 semillas por unidad experimental (semibandeja) donde se aplicaron los tratamientos cada 15 días y, 21 días luego de la última aplicación, se midieron las variables: peso seco (PSV), largo (LV) y diámetro de vástago (DV), peso total de hoja (PTH) y peso seco de raíz (PSR). El análisis empleado fue correlaciones de Pearson, componentes principales (ACP), considerando como unidades las plantas, y modelos lineales mixtos con la prueba Di Rienzo-Guzman-Casanoves (DGC). Se utilizó el software InfoStat con su interfaz con R con $\alpha=0,05$. No se encontraron correlaciones entre las variables PSR-PSV (valor- $p=0,8175$) y PSR-LV (valor- $p=0,3202$), el resto presentaron significancia. Las dos primeras componentes del ACP explicaron el 77% de la variabilidad. Sólo se detectó heterocedasticidad en PTH. Para PSR y LV sólo se diferenció el control (valor- $p=0,0283$ y $0,0244$), presentando menor media. En PSV no hubo diferencias. En DV existieron diferencias (valor- $p=0,0102$) y a través de la prueba DGC se observaron dos grupos, uno control y aplicación mixta; y otro, foliar y radicular. Los resultados fueron similares a lo observado en el biplot de ACP. Se recomienda el sistema de aplicación foliar por facilidad aplicación y se concluye que el uso de bioestimulante a base de algas favorece el crecimiento y desarrollo de las plantas de alhelí cultivadas en bandejas multiceldas, y permite obtener plantines de alta calidad comercial.

Palabras clave: **bioestimulante, análisis de componentes principales, modelos lineales mixtos, alhelí, especie floral.**

ANÁLISIS MULTIVARIADO PARA LA EVALUACION DE FUNGICIDAS DE USO AGRÍCOLA E IDENTIFICACIÓN DE NUEVOS COMPUESTOS

MARIA LILIANA MIRANDA^{1,2}, DANTE STRATICO^{1,2}, VICTORIA RICHMOND^{1,3}, MARCELO OTERO⁴

¹UMYMFOR (CONICET-Universidad de Buenos Aires)

²Depto. de Química Biológica, Fac. de Ciencias Exactas y Naturales, Universidad de Buenos Aires

³Depto. de Química Orgánica, Fac. de Ciencias Exactas y Naturales, Universidad de Buenos Aires

⁴IFIBA (CONICET-UBA), Depto. de Física, Fac. de Ciencias Exactas y Naturales, Universidad de Buenos Aires

Aires
mlmiranda@qo.fcen.uba.ar

RESUMEN

Las enfermedades de origen fúngico representan pérdidas de cultivos alimentarios entre el 10 y 15%. Adicionalmente, la aparición de patógenos resistentes a los fungicidas comerciales se ha acelerado en las últimas décadas, es por ello que la búsqueda de nuevos fungicidas continúa siendo una necesidad en el campo de investigación. En este sentido, el uso de herramientas computacionales y estadísticas para la identificación de nuevos compuestos presenta las mismas ventajas que aquellas demostradas en el desarrollo de fármacos: reducción de tiempo y costos. Nos hemos propuesto así identificar a partir de una base de datos (*DrugBank*), aquellos compuestos que puedan actuar como fungicidas frente a algunos patógenos que afectan los principales cultivos en Argentina. Con el objetivo de estudiar las propiedades de los fungicidas agrícolas, se evaluaron 210 compuestos reportados por “*Fungicide Resistance Action Committee*” (FRAC), clasificados por su mecanismo de acción y blanco molecular. Para el estudio de ambas bases de datos se calcularon 19 descriptores moleculares utilizando el programa ChemAxom y se analizaron mediante métodos de reducción de la dimensionalidad como el Análisis de Componentes Principales (PCA), ejecutado con el paquete R, así como por dos métodos no lineales: Umap y t-SNE. Este análisis nos permitió agrupar algunos de los fungicidas comerciales de acuerdo con sus blancos moleculares con los descriptores usados. En el caso del PCA, adicionalmente se proyectó la base de *DrugBank* sobre el espacio descrito por los fungicidas comerciales y esto permitió identificar compuestos candidatos que potencialmente podrían actuar sobre los mismos blancos. Este trabajo se complementó con técnicas de clusterización para encontrar compuestos con posible acción en más de un blanco, *docking* molecular (con Autodock Vina) y finalmente para evaluar la estabilidad de la interacción macromolécula-compuestos candidatos, se realizaron simulaciones de dinámica molecular, empleando el programa Amber.

Palabras clave: **Fungicidas – Descriptores moleculares – Reducción de dimensionalidad.**

ANÁLISIS MULTIVARIADO PARA LA EVALUACIÓN DE LA FUSIBILIDAD DE LAS CENIZAS DE BAGAZO Y RAC DE CAÑA DE AZÚCAR DE TUCUMÁN

PEÑA MALAVERA A. N.¹; ZAMORA RUEDA G. DEL H.²; GUTIERREZ C. E.²; FEIJÓO E.², RUÍZ, R. M.²; y GOLATO M.A.²

¹ Instituto de Tecnología Agroindustrial del Noroeste Argentino (ITANOA), Estación Experimental Agroindustrial Obispo Colombres (EEAOC) - CONICET (CONICET), CCT NOA Sur

² Estación Experimental Agroindustrial Obispo Colombres (EEAOC)

andreapema@eeaoc.org.ar

RESUMEN

El bagazo y los residuos de cosecha de caña de azúcar (RAC) son biomásas con gran potencial para la generación de energía. Sin embargo, su aprovechamiento presenta desafíos técnicos y altos costos, principalmente debido a problemas de combustión y transporte. Las cenizas generadas contienen componentes inorgánicos como silicio, calcio, magnesio, sodio, potasio, aluminio, hierro, titanio, fósforo y azufre, que afectan las temperaturas de fusión de las cenizas (DT: inicio de fusión, ST: ablandamiento, HT: hemiesfera, FT: fusión) según la norma ASTM1857. Este estudio tuvo como objetivo analizar la composición química de las cenizas de estas biomásas y su relación con las temperaturas de fusión. Se recolectaron y procesaron 30 muestras de bagazo y 30 de RAC de la variedad LCP85-384 durante las zafras 2016 a 2019 en diferentes ingenios y campos cañeros de Tucumán. Posteriormente, se realizó un análisis descriptivo de los componentes mayoritarios de las cenizas y un análisis de componentes principales (ACP) para identificar la asociación entre las variables químicas y las distintas temperaturas de fusión. El ACP reveló patrones importantes en la composición química y su impacto en la fusibilidad de las cenizas. En el caso del bagazo bajo atmósfera oxidante (AO), se observó una asociación negativa entre %SiO₂ y DT, mientras que %K₂O mostró poca influencia. En atmósfera reductora (AR), %Fe₂O₃, %CaO y %MgO se asoció positivamente con DT, y %K₂O influyó en ST, HT y FT. En las cenizas de RAC bajo AO, %Al₂O₃ y %SiO₂ presentaron correlaciones positivas con DT, mientras que %CaO fue negativo. Bajo AR, %CaO, %SO₃ y %K₂O mostraron influencias negativas en las temperaturas de fusión, destacando la importancia de los componentes inorgánicos en el comportamiento de las cenizas.

Palabras clave: **biomasa, análisis de componentes principales, composición química, combustión**

ESTUDIO EXPLORATORIO SOBRE LAS PREFERENCIAS DEL CONSUMIDOR DE CARNE DE CORDERO

ROVEGNO, MS¹, DAMICO, AB¹, DI PASQUALE, J², MICCOLI, F¹³, ABBIATI, NN¹

¹ Instituto de Investigación sobre la Producción Agropecuaria, Ambiente y Salud (IIPAAS), Facultad de Ciencias Agrarias, Universidad Nacional de Lomas de Zamora

² Dipartimento di Medicina Veterinaria, Università degli Studi di Teramo

³ Facultad de Agronomía, Universidad de Buenos Aires

mrovegno@agrarias.unlz.edu.ar

RESUMEN

Mundialmente se prevé un aumento en el consumo de carnes, con una expectativa de crecimiento particular para la carne ovina. El objetivo del estudio fue explorar las percepciones sobre la carne ovina, identificando sus atributos y las preferencias de los consumidores. Durante un evento social, se realizó una degustación de pierna de cordero cruza Texel a la parrilla, y se entregó una encuesta con cuatro secciones (información demográfica, atributos, consumo y degustación de carne de cordero) con preguntas cerradas, semicerradas y con respuestas de valoración en una escala de 0 a 10. Se recogieron 82 cuestionarios. Para el análisis se emplearon estadísticas descriptivas y técnicas multivariadas (análisis de correlación y componentes principales, ACP, para las respuestas de valoración y correspondencias múltiples, ACM, para el resto). Se utilizó el software Navure, con $\alpha=0.05$. El análisis descriptivo mostró que el 71% de las mujeres y el 75% de los hombres consumen carne de cordero ocasionalmente, preferentemente en reuniones sociales y preparada a la parrilla. Se encontraron correlaciones significativas entre el sabor y la cantidad de grasa (valor- $p<0.0001$), así como entre la grasa y la intensidad del olor (valor- $p=0.0003$). El biplot del ACP indicó que los hombres de edad intermedia perciben la carne de cordero más costosa que la vacuna, mientras que las mujeres adultas la valoran principalmente por su sabor. Por último, el gráfico del ACM reveló que, para los aspectos positivos, los jóvenes destacaron la ternura de la carne, mientras que los adultos intermedios y mayores su sabor característico. En los aspectos negativos todos los indagados señalaron la sequedad, la dificultad en la preparación y el acceso limitado al producto. Las técnicas multivariadas empleadas permitieron destacar la importancia de comprender las preferencias y percepciones del consumidor para desarrollar estrategias que impulsen el consumo de carne ovina en el país.

Palabras clave: **comportamiento del consumidor, carne ovina, análisis de componentes principales, análisis de correspondencia múltiple**

COMPARACIÓN DE LA PERFORMANCE DE MODELOS MORFOMÉTRICOS PREDICTORES DEL SEXO EN *POMACEA* (CAENOGASTROPODA: AMPULLARIIDAE)

RODRIGO G. VILLARREAL¹, EDUARDO M. SOTO¹, IGNACIO M. SOTO¹, SERGIO E. MIQUEL², ANDRÉS
O. PORTA^{1,3,4}

¹ *Departamento de Ecología, Genética y Evolución, Instituto de Ecología, Genética y Evolución (IEGEB), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires-CONICET*

² *División Invertebrados, Museo Argentino de Ciencias Naturales Bernardino Rivadavia, Buenos Aires*

³ *División Aracnología, Museo Argentino de Ciencias Naturales Bernardino Rivadavia, Buenos Aires*

⁴ *Departamento de Ciencias Exactas, Universidad Nacional del Oeste, San Antonio de Padua, Buenos Aires*

rodrivillarreal29@gmail.com

RESUMEN

Las técnicas de morfometría geométrica son de uso frecuente en biología pues permiten describir y analizar la variación de las formas de los seres vivos. Una metodología empleada frecuentemente para describir contornos cerrados (*e.g.*, hojas, valvas) es el análisis de contornos de Fourier (ACF). Este enfoque se basa en modelar la forma de un contorno por medio de una serie de funciones armónicas, cuyos coeficientes pueden ser empleados en un análisis multivariado para estudiar las variaciones en la conformación y el tamaño de individuos a comparar. Con el propósito de contrastar la efectividad de este método con otros basados en medidas lineales, se elaboraron dos modelos con datos de primera mano que permiten determinar el sexo de individuos de las especies de gasterópodos dulciacuícolas *Pomacea canaliculata* (Lamarck, 1822) y *Pomacea scalaris* (d'Orbigny, 1835) utilizando la forma y las dimensiones del opérculo. Uno de ellos utiliza morfometría clásica (con medidas lineales), mientras que el otro emplea morfometría geométrica a partir de un ACF. En una primera etapa se desarrollaron y entrenaron modelos que permiten estimar la verosimilitud del sexo de un individuo utilizando cada tipo de datos, empleando para esto una fracción del total de la muestra. En la segunda etapa, de testeo, se evaluó la efectividad de cada modelo obtenido en la etapa anterior y de otro desarrollado por N. J. Cazzaniga en 1990 (sólo aplicable a *P. canaliculata*) para estimar el sexo de los individuos de la fracción restante de los muestreos. Para *P. canaliculata*, los resultados permiten concluir que los dos modelos desarrollados poseen una eficacia similar y superior al propuesto por Cazzaniga. En el caso de *P. scalaris*, el modelo basado en coeficientes de Fourier presenta una capacidad predictora levemente superior, aunque su eficiencia depende del sexo, con mejor rendimiento para clasificar hembras.

Palabras clave: **Morfometría, Clasificación, Modelización, Verosimilitud.**

AGRUPAMIENTOS BASADOS EN MIXTURAS FINITAS GAUSSIANAS: IMPACTO DE LA DIMENSIONALIDAD Y EL TAMAÑO DE MUESTRA EN DATOS SIMULADOS DE MORFOMETRÍA GEOMÉTRICA

JUAN ESTEBAN VRDOLJAK^{1,2}, FEDERICO ARES^{2,3}, SANTIAGO BENITEZ-VIEYRA⁴

¹ Instituto Patagónico para el Estudio de los Ecosistemas Continentales, CONICET, Puerto Madryn, Chubut

² Universidad Nacional de la Patagonia, Puerto Madryn, Chubut

³ Instituto Patagónico de Geología y Paleontología, CONICET, Puerto Madryn, Chubut

⁴ Instituto Multidisciplinario de Biología Vegetal, Universidad Nacional de Córdoba, CONICET, Córdoba

juan.vrdoljak@gmail.com

RESUMEN

Las técnicas de agrupamientos basados en modelos de mixturas finitas gaussianas son de las técnicas para estudiar regularidades más popular para datos continuos mientras que la morfometría geométrica es una herramienta útil para cuantificar formas de manera continua y muy precisa. Estudiar regularidades en un contexto de alta dimensionalidad como el que presenta la morfometría geométrica es un impedimento para los métodos de agrupamiento basados en modelos, por lo que suele reducirse el conjunto de datos a unas pocas dimensiones. El objetivo del presente trabajo es poner a prueba el desempeño de los agrupamientos basados en modelos de mixturas finitas gaussianas frente a distintas dimensiones originales y tamaños de muestra con datos simulados de morfometría geométrica. Para ello, simulamos datos con distinta cantidad de dimensiones (20, 50, 100, 200, 400, 800) y tamaño de muestra (10, 50, 100, 200, 400, 1000) alrededor de dos medias poblacionales. A cada conjunto de datos se le redujo la dimensionalidad mediante Análisis de Componentes Principales, reteniendo no menos de tres dimensiones, no más de diez ni más del 90% de la variación total. En cada simulación se estimó el desempeño a partir de la cantidad de agrupamientos (k) que mejor ajustaron a los datos y el índice de Rand (ajustado) a partir de la clasificación obtenida para $k=2$. Se obtuvo un muy bajo desempeño con diez muestras indistintamente de la cantidad de dimensiones. Asimismo, los valores extremos de dimensiones (20, 400 y 800) mostraron un desempeño moderado a bajo. Finalmente, recomendamos utilizar una cantidad moderada de información (entre 50 y 200 dimensiones) previo a la reducción de dimensionalidad, en conjunto con un tamaño de muestra no menor a 50, para la obtención de resultados que representen la información de los datos de morfometría geométrica mediante agrupamientos basados en modelos de mixturas finitas gaussianas.

Palabras clave: ***análisis de agrupamiento, alta dimensionalidad, modelos de mixturas finitas gaussianas, morfometría geométrica, simulación de datos.***

ANÁLISIS DE MICROBIOMA RUMINAL DE CORDEROS TEXEL EN ENGORDE A CORRAL: EFECTO DE LA INCLUSIÓN DE LINO Y DEL TIPO DE MUESTREO RUMINAL

MICCOLI, F. E.^{1,5}, CENTURION, N.D.^{2,1}, FERREYRA, S.^{3,1}, GALARZA, R.⁴, IORIO, J.⁵, JULIANO, N.⁵, WEHRENDT, D.⁶, GUERRERO, L. D.⁶.

¹Facultad de Ciencias Agrarias, Universidad Nacional de Lomas de Zamora

²CITNOBA (CONICET- UNNOBA- UNSAdA)

³Universidad Nacional de San Antonio de Areco (UNSAdA)

⁴INTA Cuenca del Salado

⁵Facultad de Agronomía, Universidad de Buenos Aires

⁶Instituto de Investigaciones en Ingeniería Genética y Biología Molecular "Dr. Héctor N. Torres", CONICET

fmiccoli@agrarias.unlz.edu.ar

RESUMEN

Se analizó el microbioma ruminal de corderos de feed lot bajo dos dietas y dos técnicas de muestreo de contenido ruminal. Corderos Texel ($n=26$, 18.6 ± 1.06 kg de peso; 43.23 ± 4.55 d de edad) fueron asignados a 2 dietas, en un diseño completamente aleatorizado: 1) CON (concentrado sin semilla de lino) y 2) LIN (concentrado con semilla de lino al 6%). El engorde duró 6 semanas. Las dietas eran isoprotéicas e isoenergéticas (17.8% Proteína, 2.72 Mcal energía metabolizable/Kg), suministradas al 5% del peso (NRC, 2007). Se seleccionaron 6 corderos por tratamiento y se muestrearon con 2 técnicas: 1) Oro-ruminal *in vivo* (IT) y 2) Directo de rumen a la faena (RD). Luego de la extracción del ADN, se analizó la región V4 del gen 16S rRNA en bacteria y archaea y se secuenció por plataforma NovaSeq PE250 (Novogene Corporation Inc., China). El análisis de las secuencias se realizó con el paquete de R 'DADA2'. Después de calcular las abundancias a nivel de filios, los filios con menos de 10 secuencias en todas las muestras sumadas fueron descartados y las muestras se refaccionaron al 90% de la muestra menor (rarefy_even_depth) con el paquete "Phyloseq". Las diferencias dentro y entre tratamientos se calcularon mediante un análisis de varianza usando distancias de matrices (distancia Bray-Curtis) sobre las abundancias relativas a partir de las muestras rarefaccionadas (adonis2 del paquete "Vegan"). No hubo efecto de la dieta sobre el microbioma ruminal a nivel Phylum ($P = 0.255$) no obstante, se detectaron diferencias entre técnicas de muestreo ($P = 0.001$); no hubo interacción Dieta x Muestreo ($P = 0.697$). Se identificaron 20 Phylum, 5 difirieron entre IT and RD: Firmicutes (63.24%; 74.3%), Bacteroidota (26.49%; 6.27%), Proteobacteria (2.70%; 10.64%) Synergistota (0.073%; 0.52%), Cyanobacteria (0.054%; 1.42%). A nivel Género, 68 taxa resultaron diferentes entre técnicas ($P < 0.05$). Los más abundantes en IT y RD fueron: *Ruminococcus* (5.11; 11.61%), UCG-005 (10.66, 0.59%), *Lachnospiraceae* NK3A20 group (1.97, 4.25%), *Blautia* (4.14, 0.30%), *Bacteroides* (4.16, 0.23%), NK4A214 group (0.67, 3.71%), *Prevotella* (0.78, 3.38%), *Succinivibrionaceae* UCG-002 (0.074, 3.57%), *Turicibacter* (2.99, 0.25%), *Prevotella_9* (2.42, 0.12%). El microbioma ruminal en corderos de feed lot difirió según la técnica de muestreo con lo cual debería tenerse en cuenta este aspecto en estudios nutricionales ligados a metagenómica.

Palabras clave: **microbioma ruminal, corderos, feed lot, método de muestreo.**

Ciudad de Buenos Aires, octubre 2024

EVALUACIÓN DE DETERMINANTES DE LA MULTIFUNCIONALIDAD ECOSISTÉMICA DEL SUELO POR MEDIO DE SEM-PLS

MARTÍN NICOLÁS RIBERO¹, MARÍA ROMINA SCHIAFFINO^{2,3}, JULIETA FILLOY¹

¹ *Laboratorio de Ecología de Comunidades y Macroecología, Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires; IEGEBA, UBA - CONICET*

² *Departamento de Ciencias Básicas y Experimentales, Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA)*

³ *Centro de Investigaciones y Transferencia del Noroeste de la Provincia de Buenos Aires (CITNOBA) – UNNOBA-UNSAaA CONICET.*

martinribero@gmail.com/mrribero@ege.fcen.uba.ar

RESUMEN

La multifuncionalidad ecosistémica del suelo (MES) se asocia tanto a variables abióticas del clima y el suelo, como a la riqueza de las comunidades microbianas. A su vez, la actividad humana, dado que podría influir indirectamente a la MES al modificar las características ambientales y la biodiversidad. Las variables relacionadas con la MES suelen ser interdependientes, por lo que es necesario estudiar las influencias directas e indirectas entre ellas y sobre la multifuncionalidad. Los modelos de ecuaciones estructurales por medio de cuadrados mínimos parciales (SEM-pls) permiten proponer múltiples variables medidas a campo representando constructos teóricos con el objetivo de eliminar aquellas correlacionadas e identificar aquellas que mejor soportan el modelo planteado. El objetivo del trabajo fue evaluar la utilidad del enfoque SEM-pls para estudiar las relaciones entre la MES, las variables ambientales, la diversidad microbiana y el uso de la tierra, a lo largo de un gradiente climático de 200 km en la provincia de Entre Ríos. En el SEM-pls inicial se incluyeron variables relacionadas con el clima (temperatura y precipitación) y con el suelo (parámetros fisicoquímicos). La biodiversidad del suelo fue representada por la diversidad taxonómica y filogenética de bacterias y hongos. El uso de la tierra fue representado por variables de conversión de la tierra e intensificación. El proceso de ajuste del SEM-pls requirió eliminar múltiples variables para llegar al mejor modelo, el cual presentó una asociación significativa entre el clima y la MES, y negativa entre el uso de la tierra y la biodiversidad ($p < 0.05$). Sin embargo, no se detectó una asociación significativa entre la biodiversidad y la MES. En el sistema estudiado, el uso de SEM-pls permitió identificar las principales asociaciones entre las variables predictoras y la MES, a pesar de la complejidad y las interacciones entre los factores evaluados.

Palabras clave: **degradación de carbono, metagenómica, pastizales, plantaciones forestales.**



Aprendizaje Automático

ENFOQUE DE APRENDIZAJE AUTOMÁTICO PARA LA SELECCIÓN DE VARIABLES APLICADO A XENOMETABOLOMICA

PABLO MARTÍN DEMETRIO¹, CAROLINA SALGADO COSTA¹, FLORENCIA BAHL MARÍA¹, GUILLERMO SEBASTIÁN NATALE¹, SARA RODRIGUEZ-MOZAZ² & LÚCIA HELENA SANTOS²

¹Centro de Investigaciones del Medioambiente (CIM-CONICET-Universidad Nacional de La Plata)

²Instituto Catalán de Investigación del Agua ICRA-CERCA)

pablo.demetrio@quimica.unlp.edu.ar

RESUMEN

Los datos del presente trabajo forman parte de un estudio xenometabólico no-dirigido realizado con plasma de anuros adultos de *Leptodactylus luctator* (N = 26) de dos sitios de Buenos Aires (Argentina) con diferentes niveles de perturbación antrópica (sitio referencia y sitio contaminado). El conjunto de datos consta de 2393 variables que son los analitos endógenos y exógenos identificados en plasma por LC-HRMS, habiéndose computado como datos censurados por izquierda a más del 10% de los valores. El objetivo del presente trabajo fue seleccionar un subconjunto de entidades con alto poder predictivo para la clasificación de ambos sitios. Tras un pre-procesamiento que incluyó centrado y escalado, eliminación de variables altamente correlacionadas o con varianza próxima a cero e imputación de valores aleatorios en el intervalo [0-señal mínima], se retuvieron 1920 variables cuantitativas. Se consideraron modelos de clasificación lineales, no-lineales y basados en árboles, reteniendo dentro de cada grupo los métodos de mayor rendimiento (todos con Accuracy > 0.9): glmnet, svm y random forest, respectivamente. Se trabajó con una partición 70% training + 30% test y para cada modelo ajustado, se evaluó el rendimiento por remuestreo (10 fold-CV x 5 repeticiones). Se seleccionaron las 100 variables más importantes de cada modelo y se combinaron, resultando en un subconjunto final de 164 entidades. Se generaron representaciones gráficas de los resultados mediante biplot y heatmap para su interpretación. Los análisis y gráficos fueron implementados en R. La utilización de distintas metodologías de selección de variables modelo-dependiente en paralelo y su posterior combinación, permitió obtener un alto poder predictivo y reducir considerablemente las variables (>90%). Las entidades seleccionadas serán evaluadas para determinar su origen (endógeno/exógeno) y explorar su potencial como biomarcadores de efecto para el diagnóstico de sitios perturbados utilizando anuros como objeto de estudio.

Palabras clave: **clasificación, ómicas, predicción, biomarcadores, machine learning, ecotoxicología.**

ESTADO DE AVANCE DEL ENSAYO PILOTO DE GENOMAS DE ALTA Y BAJA COBERTURA DEL PROGRAMA POBLAR

FIGUEROA MARCELO ISIDRO^{1,2}, LUISI PIERRE^{1,3}, SORIA DANIEL^{1,4}, FERNÁNDEZ EZEQUIEL⁵, BRAVI CLAUDIO¹, BAILLIET GRACIELA¹, ARGÜELLES CARINA¹, MIRETTI MARCOS¹, DIPIERRI JOSÉ EDGARDO¹, ALFARO EMMA¹, RAMALLO VIRGINIA¹, GONZALES-JOSÉ ROLANDO¹, J. SÁNCHEZ-LORIA⁶, T. POKLEPOVICH⁶ y DOPAZO HERNÁN^{1,4}.

¹ Programa de Referencia y Biobanco de la Población Argentina (PoblAR), (SPyPC-MINCYT).

² Instituto de Ecorregiones Andinas (INECOA-CONICET)

³ Instituto de Antropología de Córdoba (IDACOR-CONICET)

⁴ Instituto de Ecología, Genética y Evolución (IEGEB), Universidad de Buenos Aires-CONICET

⁵ Fundación Actyon

⁶ Administración Nacional de Laboratorios e Institutos de Salud "Dr. Carlos Malbrán" (ANLIS Malbrán)

mfigueroa@conicet.gov.ar

RESUMEN

PoblAR (Programa de Referencia y Biobanco Genómico de la Población Argentina) es el primer biobanco nacional de datos genéticos, hábitos alimentarios, y características fenotípicas y sociales de nuestro país. Para avanzar en la definición de los métodos genómicos, estadísticos y bioinformáticos que utilizará el programa, nos propusimos completar un ensayo piloto inicial utilizando 100 muestras enriquecidas con ancestría nativa americana. El objetivo de este piloto fue optimizar y validar el flujo de trabajo derivado de los datos genómicos utilizando genomas completos secuenciados a alta (HC) y baja cobertura (LC); evaluando, a través de los genomas HC, la precisión y validez de los resultados de imputación en LC. Las 100 muestras fueron secuenciadas en el Instituto de Genómica y Bioinformática del ANLIS-Malbrán. El procesamiento bioinformático se realizó con scripts propios y software libre en el clúster Mulatona del CCAD (UNC). Para contrastar los resultados de los análisis de HC utilizamos las herramientas de Illumina Connected Analytics (ICA-Dragen-Germline). La anotación funcional en muestras HC resultó en una mediana de 4,64 millones de variantes/genoma (3,55 M de SNPs y 1,1 M de INDELS por genoma). El porcentaje de genotipos no coincidentes entre ambas estrategias bioinformáticas fue < 1% en media ($\approx 34,7K$ SNPs). En genomas LC el número de genotipos imputados aumentó de forma significativa con el aumento de la cobertura y extensión del genoma secuenciado. En categorías de variantes de frecuencias < 1% este error nunca fue superior al 10% en media. Finalmente, los estudios de PCA y proporciones de ancestría entre genomas HC y LC no mostraron diferencias que impidan la correcta interpretación de los resultados. Estas conclusiones subrayan la posibilidad de llevar adelante un programa de reclutamiento, secuenciación y análisis efectivo de datos genómicos a costos razonables para países como el nuestro donde la inversión en I+D+i es cada vez menor.

Palabras clave: **PoblAR – Genómica Poblacional - Bioinformática.**

Ciudad de Buenos Aires, octubre 2024

ALGORITMOS DE APRENDIZAJE AUTOMÁTICO EN LA PREDICCIÓN DEL COMPORTAMIENTO A ENFERMEDADES Y RENDIMIENTO EN DOS CULTIVOS

SEBASTIAN FILIPIGH¹, FRANCO SUAREZ¹, JULIANA IGLESIAS³, CECILIA BRUNO^{1,2}

¹*Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba*

²*Grupo vinculado de Estadística y Biometría Unidad de Fitopatología y Modelización Agrícola (UFyMA-INTA-CONICET), Córdoba*

³*INTA-Pergamino-UNNOBA*

sebastian.filipigh@mi.unc.edu.ar

cebruno@agro.unc.edu.ar

RESUMEN

La predicción genómica en plantas permite estimar características fenotípicas en individuos no evaluados. Este estudio compara la capacidad predictiva de modelos de aprendizaje automático (ML) frente a modelos lineales mixtos (MM). Para ello se emplearon dos bases de datos: una del Programa Global de Mejoramiento de Trigo (CIMMYT, México) con 599 líneas genotipadas con 11447 marcadores DArT y valores de rendimiento; y otra del INTA Pergamino (Argentina) con 63 líneas de maíz genotipadas con un chip de SNPs de 56K y fenotipadas para *Fusarium verticillioides* y *Fusarium graminearum*. Se ajustaron tres modelos de ML: 1-Bosques Aleatorios (RF), 2-Máquinas de Soporte Vectorial (SVM), y 3-Máquinas de Gradiente Recurrente (GBM) y tres MM con efecto aleatorio de genotipo y tres estructuras de varianza y covarianza: matriz de parentesco (K), matriz de aditividad genética (A) y la combinación de ambas (KA). Se realizó una validación cruzada para evaluar el comportamiento de los modelos en la predicción, se estimó la correlación (r) entre los valores observados y los predichos y se calculó el error cuadrático medio de predicción (RMSE). En el caso del trigo, el MM-A fue el de menor RMSE= 0.84 y obtuvo la mayor correlación $r=0.50$. Entre los ML, RF presentó menor RMSE=0.77 y la mayor correlación $r=0.62$. En cuanto al maíz, para FV se obtuvo RMSE=0.28 y valor de correlación $r=0.98$ y para FG el RMSE fue de 0.28 y valor de correlación $r=0.97$. Los modelos de aprendizaje automático ofrecieron buenos resultados predictivos en comparación con los modelos lineales mixtos, siendo RF el de mejor comportamiento predictivo para los caracteres agronómicos evaluados en este estudio.

Palabras clave: **correlación genética, modelos lineales mixtos, capacidad predictiva**

APLICACIÓN DEL DESCRIPTOR ESTADÍSTICO LBP PARA EL RECONOCIMIENTO AUTOMÁTICO DE TORTUGAS VERDES EN IMÁGENES

CINTIA FLORES YAPU¹, SILVINA PISTONESI^{1,2}, JORGE MARTINEZ¹, ANA GEORGINA FLESIA³ Y GABRIELA VELEZ RUBIO⁴

¹*Departamento de Matemática, Universidad Nacional del Sur*

²*Departamento de Ciencias Básicas, Facultad Regional Bahía Blanca, Universidad Tecnológica Nacional*

³*Facultad de Matemática, Astronomía y Física, Universidad Nacional de Córdoba - Centro de Investigaciones en Matemática*

⁴*Sección de Oceanografía y Ecología Marina, Instituto de Ecología y Ciencias Ambientales, Facultad de Ciencias, Universidad de la República - Asociación Civil Karumbé, Uruguay*

cintia.flores@uns.edu.ar

RESUMEN

Las técnicas de reconocimiento individual de animales basadas en fotografías son de vital importancia, ya que representan alternativas no invasivas y menos costosas en términos de material y tiempo de aplicación. Estas aprovechan los patrones físicos naturales del animal para la identificación del mismo. En el caso particular de las tortugas marinas, puede realizarse considerando el patrón de las escamas de los laterales de la cabeza, único para cada ejemplar. Los recientes desarrollos de algoritmos de reconocimiento de patrones texturales en imágenes hacen posible la automatización de este proceso de identificación. Entre ellos, se destaca la utilización del descriptor estadístico textural LBP (Local Binary Pattern) porque resume la estructura local de una imagen y es un operador robusto en términos de variaciones de escala de grises, invariante a rotaciones y de fácil implementación. Etiqueta las texturas mediante códigos binarios según las relaciones entre cada píxel de la imagen y sus vecinos. El objetivo de este trabajo es emplear este descriptor para automatizar el reconocimiento de individuos de tortuga verde a través de sus imágenes. Para el desarrollo del estudio se creó una base de 310 imágenes de la especie *Chelonia mydas* (compuesta de 10 clases con 31 imágenes cada una, obtenidas de transformaciones geométricas). Para evaluar la precisión de la identificación de tortugas, se utilizó el LBP en combinación con el clasificador 1-Nearest Neighbors (1-NN) y el logístico Multinomial (LM), mediante el método k-cross fold validation. Primeramente, se efectuó un análisis descriptivo para diferentes valores de los parámetros del descriptor; y luego, se aplicó el test no paramétrico de Wilcoxon. Los resultados experimentales demuestran que es posible reconocer individuos de tortuga verde a partir de imágenes de sus patrones de escamas faciales utilizando el LBP, alcanzando una mejor precisión de la identificación con el clasificador 1-NN en comparación al LM.

Palabras clave: **patrón textural, local binary pattern, escamas faciales, tortugas verdes**

Ciudad de Buenos Aires, octubre 2024

PREDICCIÓN DE DÍAS A MADUREZ EN SOJA (*GLYCINE MAX SP.*) A TRAVÉS DE IMÁGENES DE VEHÍCULOS AÉREOS NO TRIPULADOS

FIORE JUAN MANUEL¹, PEROTTO CECILIA³, BRUNO CECILIA^{1,2}

¹*Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba*

²*Grupo de Estadística y Biometría vinculado a la Unidad de Fitopatología y Modelización Agrícola (UFyMA) – INTA-CONICET.*

³*Centro de Investigaciones Agropecuarias-CIAP, UFyMA, INTA-CONICET*

fiorejuanm@gmail.com

RESUMEN

En los programas de mejoramiento genético vegetal, los días a madurez (DM) son utilizados para comparar variedades del mismo ciclo y seleccionar las de mejores aptitudes agronómicas. Esta variable determina el largo del ciclo del genotipo, el cual varía a lo largo de la Argentina y está altamente correlacionada con la temperatura y el fotoperiodo. La determinación de los DM en soja representa un desafío en los programas de mejoramiento al momento de seleccionar genotipos según su rendimiento. Además, son utilizados como información complementaria en la asignación de su región de adaptación. El objetivo de este trabajo es estimar los DM a través de imágenes de vehículos aéreos no tripulados tomadas sobre ensayos comparativos de rendimiento (ECR) con Redes neurales convolucionales (CNN). Se trabajó con datos provenientes de un ECR conducido en dos localidades. Se entrenó una CNN que toma como input/regresora una imagen de 20×150 píxeles. Para cada localidad se utilizaron ~26.000 imágenes. Cada imagen fue clasificada como seco y no seco según la lectura de los DM realizada en el campo. Una localidad fue utilizada para entrenar el modelo y la otra para validarlo. Los valores de DM estimados por el modelo fueron comparados con lecturas de DM realizadas en el campo y se estimó el Error Cuadrático Medio (RMSE). En el conjunto de datos de entrenamiento el RMSE fue de 2.56 DM y en el conjunto de datos validación de 2.77 DM. Las principales fuentes de error se encontraron en imágenes donde coincidieran dos surcos y donde crecían malezas o re-emergían plantas. La estimación de DM usando el algoritmo CNN resultó una herramienta efectiva proporcionando una alternativa precisa y automatizada a las mediciones manuales tradicionales. El modelo de predicción de DM a partir de CNN puede ser usada en múltiples etapas de un programa de mejoramiento genético vegetal, independientemente del número de surcos.

Palabras clave: **Redes neuronales, ensayos comparativos de rendimiento, grupos de madurez, drones.**

ELABORACIÓN DE UN MODELO DE PRONÓSTICO PARA FLORACIONES ALGALES EN EL EMBALSE SAN ROQUE, CÓRDOBA

MICAELA JUANEDA ALLENDE^{1,2}, MARCELO SMREKAR¹, PEDRO MANSUR¹, PEDRO RUIZ DÍAZ¹, ARACELI FERREYRA¹, JONATHAN MUCHIUT^{1,2}, ALBA GERMAN³, ANDRES RODRÍGUEZ^{1,2}

¹ *Laboratorio de Hidráulica, Facultad de Ciencias Exactas Físicas y Naturales, Universidad Nacional de Córdoba*

² *CONICET, Córdoba*

³ *Departamento de Geografía, Facultad de Filosofía y Humanidades, Universidad Nacional de Córdoba*

micajuaneda@mi.unc.edu.ar

RESUMEN

Desde 1970 y hasta la actualidad, se han observado eventos de floraciones algales en el embalse San Roque de la provincia de Córdoba, afectando el proceso de potabilización del agua, la salud pública y las actividades recreativas. En particular, la dominancia fitoplanctónica recae sobre las cianobacterias del género *Microcystis*, potencialmente tóxicas. El desarrollo de un modelo predictivo de concentración de clorofila-a, indicador de biomasa algal, es de gran interés para la gestión del recurso hídrico que presenta condiciones de un cuerpo de agua eutrofizado. En este trabajo se exponen resultados preliminares de un modelo de aprendizaje automático que utiliza un dataset complejo respecto a su confección y curación de los datos. En donde se optó por un algoritmo de XGBoosting y se propone una métrica novedosa para la evaluación de la aplicación del modelo en diversas clasificaciones como función de un parámetro a partir de las métricas usuales. Resulta de interés la mejora constante en los modelos de aprendizaje automático y el brindar información certera sobre las variables regresoras, por lo que se emplea una herramienta conocida como SHAP para el análisis detallado de la incidencia de cada variable en la predicción.

Palabras clave: ***Eutrofización, cianobacterias, XGBoosting, SHAP***

DETECCIÓN DE CRIADEROS POTENCIALES DE *Aedes albifasciatus* EN UNA CUENCA HÍDRICA METROPOLITANA, A PARTIR DEL ANÁLISIS DE IMÁGENES SATELITALES

LEON, MARTIN JAVIER¹; CAVIA, REGINO² ; GARZON, MAXIMILIANO¹

¹Grupo de Estudio de Mosquitos, Departamento de Ecología Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

² Laboratorio de Ecología de Poblaciones, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

martinjavierleon20@gmail.com

RESUMEN

Aedes albifasciatus es un mosquito de inundación. Sus estadios inmaduros se desarrollan principalmente en cuerpos de agua temporales formados a partir de precipitaciones (zonas anegables). Estos cuerpos suelen encontrarse en depresiones de planicies cubiertas de pastizales donde las hembras depositan sus huevos. Este estudio se centró en desarrollar un método que emplea análisis de imágenes satelitales para detectar zonas anegables que pudieran servir como criaderos. En un parque de referencia, se evaluaron diferentes índices (NDVI, NDMI, NDWI, TVI, VSWI) usando la plataforma Google Earth Engine y Análisis de Componentes Principales, comparando su variación entre zonas anegables y no anegables. Se encontró que las variaciones en el NDMI mostraban las mayores diferencias zonas anegables y no anegables, con respecto a otros índices. Se creó una composición multibanda de NDMI de distintas fechas del área de la cuenca del río Matanza-Riachuelo y se realizó una clasificación de ambientes anegables y no anegables con el algoritmo "Random Forest". El resultado fue un mapa de sitios anegables y no anegables de la cuenca. La precisión del mapa se evaluó indirectamente mediante un muestreo de mosquitos adultos capturados en 33 sitios de muestreo distribuidas en el área, donde también se midieron variables ambientales relacionadas. Mediante un Modelo Lineal Generalizado, se analizó la asociación entre las características ambientales relevadas y la abundancia de *Aedes albifasciatus* capturados. El tipo de cobertura alrededor del sitio, la zona (urbana, periurbana o rural) y dos interacciones entre variables: 1) Proporción de área anegable en 500m alrededor de cada sitio con la distancia a un camino pavimentado, y 2) Distancia a un camino pavimentado con signos de lluvia reciente, explicaron la variación de la abundancia de *Ae. albifasciatus*. Los resultados evidencian que la clasificación supervisada de variaciones temporales del NDMI permitiría identificar los criaderos del mosquito.

Palabras clave: *mosquito, SIG, teledetección, zonas anegables, índice espectral.*



DETECCIÓN DE DIFERENCIABILIDAD EN LA VOZ UTILIZANDO TÉCNICAS DE RESAMPLING SOBRE AUDIOS GRABADOS CON SMARTPHONES

RUBÉN ALFREDO MENDOZA¹

¹ *Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba*

ruben.mendoza@mi.unc.edu.ar

RESUMEN

El objetivo de este estudio fue evaluar si es posible detectar diferencias significativas en la voz de hombres y mujeres a partir de audios grabados con smartphones. Para la recolección de datos, se desarrolló un bot de Telegram que guió a los usuarios en la grabación de seis audios en smartphones de distintas marcas y modelos: cinco consistieron en la pronunciación de las vocales y uno contando del 0 al 9. Se extrajeron un total de 82 características acústicas de estos audios y se compararon las medias de dichas características entre ambos sexos mediante pruebas de permutación, con un nivel de significancia (α) de 0,05. Para corregir los p-valores por múltiples comparaciones, se aplicó la corrección de Hochberg. Como resultado, se encontraron diferencias estadísticamente significativas en 2 a 13 características acústicas, dependiendo del audio analizado. Posteriormente, las características fueron reducidas mediante análisis de componentes principales (PCA). Con estas componentes principales, se entrenaron seis modelos de regresión logística, uno para cada tipo de audio, con el objetivo de identificar las características más relevantes a la hora de predecir el sexo del hablante. Las probabilidades de pertenencia a cada categoría, generadas por cada modelo, fueron utilizadas para calcular una probabilidad total, logrando así una clasificación más robusta. Además, se implementó un segundo bot de Telegram que permite a los usuarios enviar nuevos audios y recibir la probabilidad de ser clasificados como hombre o mujer en función de las predicciones de los modelos. Este trabajo no sólo valida la capacidad para detectar diferencias de sexo en la voz utilizando diversos smartphones, sino que también sienta las bases de una metodología integral, aplicable en futuros estudios para la detección de trastornos de salud a través de la voz.

Palabras clave: ***Python, Caracterización acústica, Estadística no Paramétrica, Análisis de Componentes Principales, Regresión Logística***



MORFOMETRÍA GEOMÉTRICA Y REDES NEURONALES CONVOLUCIONALES: OPTIMIZACIÓN DE LA IDENTIFICACIÓN DE SUBESPECIES DE *APIS MELLIFERA*

GUERRERO, AURELIANO¹; PORRINI, LEONARDO^{2,3}; BRADESCO, CONSTANZA^{2,3}; CHAPARRO, MAURO^{1,3}

¹*Centro Marplatense de Investigaciones Matemáticas (CEMIM), Departamento de Matemáticas, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata.*

²*Instituto de Investigaciones en Producción, Sanidad y Ambiente (IIPROSAM), Universidad Nacional de Mar del Plata (UNMdP). Centro de asociación simple CIC-PBA. Mar del Plata, Buenos Aires*

³*CONICET (CONICET)*

aguerrer@mdp.edu.ar

RESUMEN

La abeja melífera (*Apis mellifera*) presenta un amplio rango de distribución mundial con poblaciones que varían tanto en la forma, así como también en sus patrones de conducta y dinámicas poblacionales asociadas a la productividad. Por medio del análisis de caracteres morfológicos se han descrito cerca de 25 subespecies, pertenecientes a 5 linajes genéticos distintos. El análisis cuantitativo de la variación fenotípica y sus relaciones con la variación genética y ambiental, ayuda enormemente a entender la naturaleza y la acción de los factores que alteran su distribución. La morfometría geométrica utiliza ciertos puntos llamados “puntos homólogos” específicos, localizados sobre una estructura biológica o una imagen de ella. En *A. mellifera*, se utilizan un conjunto de puntos de intersección en las ramificaciones venosas de las alas delanteras. A partir de estas se obtienen imágenes en las que vamos a referenciar los puntos homólogos que seleccionamos. El objetivo de este trabajo es desarrollar una metodología novedosa y ágil que permita la identificación de 19 puntos homólogos y describir linealmente las ramificaciones venosas con el fin de clasificar las colonias de abejas a nivel de subespecies de *A. mellifera*. Para esto, se entrenaron a partir de un banco de 7000 imágenes de alas, redes neuronales convolucionales tales como Florense2 y Yolov8-pose. El entrenamiento y la optimización de ambas redes utilizadas permitirán clasificar correctamente poblaciones de abejas manejadas en diferentes regiones de Argentina. El uso de morfometría alar representa una novedosa herramienta que pretende ser aplicada a la selección de material vivo apícola. La implementación de estas tecnologías avanzadas para el procesamiento de imágenes, estudio estadístico de las estructuras geométricas resultantes y análisis de componentes principales para establecer pertenencias a grupos resultará en una mejora significativa en el proceso de selección de abejas reinas, beneficiando a la industria apícola nacional con abejas más resistentes y productivas.

Palabras clave: ***Deep Learning, Clustering, Componentes Principales***



CHAT GPT: ALIADO O ENEMIGO PARA EL ANÁLISIS ESTADÍSTICO DE DATOS

PEÑA MALAVERA A. N.¹, MUÑOZ MORALES D.²

¹ *Instituto de Tecnología Agroindustrial del Noroeste Argentino (ITANOA), Estación Experimental Agroindustrial Obispo Colombres (EEAOC) - CONICET (CONICET), CCT NOA Sur*

² *Ministerio de Seguridad de Tucumán*

andrapema@eeaoc.org.ar

RESUMEN

En las instituciones tanto públicas como privadas el rol del científico de datos se va gestando en ámbitos muy diversos, desde el análisis de experimentos, procesos industriales, pasando por análisis menos estructurados como encuestas, muestreos, etc. Así mismo estos pueden provenir de diversas áreas del conocimiento en las cuales los métodos estadísticos y los distintos software están adaptados a ellas. El chatGPT y la Inteligencia artificial (IA) en general han surgido como una novedosa herramienta no solo de escritura de textos, revisión bibliográfica sino que los más osados buscan un colaborador para el análisis estadístico de sus datos. En este trabajo hicimos consultas aleatorias al ChatGPT en análisis básicos y específicos de distintas áreas. Uno de las consultas evidenció que ChatGPT te ayuda a calcular el promedio de una variable categórica codificada en números, i.e. la variable "color" codificada con un valor entero a cada uno de los colores primarios, si el usuario no conoce la naturaleza de la variable "color" puede tomar el resultado como una verdad absoluta. Ejemplos más específicos en genética, análisis de fraude, psicometría entre otros fueron consultados para su análisis. Algunos de los errores más comunes fueron: número de alelos totales y efectivos mayores a niveles de ploidia, correlaciones fuera de rango, medias de variables categóricas etc. Como conclusiones, para realizar un análisis estadístico certero es necesario preguntar varias veces al ChatGPT sobre la misma cuestión, darle opciones de cambiar su respuesta haciendo notar algún punto crítico y revisar cada uno de los resultados que se van generando. El usuario de ChatGPT debe tener claros los conceptos estadísticos básicos, debe entender el funcionamiento del software o lenguaje de programación a usar y analizar a conciencia cada uno de los resultados. Siendo así muy importante la capacitación continua en estadística básica y avanzada.

Palabras clave: **Inteligencia artificial, Naturaleza de la variable, variable cualitativa, variable cuantitativa**



PREDICCIÓN DE RENDIMIENTOS AGRÍCOLAS EN CORRIENTES: INTEGRACIÓN DE DATOS MEDIANTE MODELOS DE APRENDIZAJE AUTOMÁTICO

SCREPNIK CLAUDIA¹, ZAMUDIO EDUARDO², GIMENEZ LAURA³

¹Facultad de Ciencias Agrarias, Universidad Nacional del Nordeste

² Universidad Nacional de Misiones

³Facultad de Ciencias Agrarias, Universidad Nacional del Nordeste

cscrepnik@comunidad.unne.edu.ar

RESUMEN

El incremento en la demanda de alimentos posiciona a la agricultura en un lugar de privilegio. Sin embargo, el sector agrícola enfrenta importantes desafíos debido al cambio climático y al uso intensivo de agroquímicos, lo que puede comprometer la seguridad alimentaria global. La provincia de Corrientes, representa aproximadamente el 21,8% de la superficie nacional de cultivo de naranjas y mandarinas. La estimación de la producción es necesaria para optimizar los resultados. La capacidad para prever con precisión los rendimientos de los cultivos constituye una herramienta para abordar estos desafíos y promover prácticas agrícolas sostenibles. El objetivo de la investigación fue desarrollar un sistema de predicción de rendimientos agrícolas que utiliza datos meteorológicos anuales, y datos históricos de producción, aplicando técnicas de aprendizaje automático. Los datos fueron suministrados por el Servicio Meteorológico Nacional, Ministerio de la Producción de Corrientes para las campañas de 1969/70 a 1996/97. Se implementaron procedimientos para la recolección, limpieza y mejora de datos, y se entrenaron y evaluaron tres modelos de aprendizaje automático: Gradient Boosting, K-Nearest Neighbors y Regresión Logística. Para optimizar el rendimiento de los modelos y minimizar el riesgo de sobreajuste, se utilizó GridSearchCV para ajustar los hiperparámetros a través de la validación cruzada K-Fold. Además, se analizó la correlación entre los rendimientos proyectados y reales de los cultivos, buscando identificar las condiciones meteorológicas óptimas. El modelo de Gradient Boosting demostró un rendimiento adecuado, con un coeficiente de determinación (R^2) de aproximadamente 75%, indicando una acertada capacidad para predecir los rendimientos. Estos hallazgos abren oportunidades para implementar métodos basados en datos en la agricultura sostenible, mejorando la gestión de recursos y avanzando hacia un futuro más seguro en términos de disponibilidad alimentaria y resiliencia frente al cambio climático.

Palabras clave: **Agricultura; predicción del rendimiento de los cultivos; aprendizaje automático**

¿CÓMO PREDIGO EL FUTURO? COMPARACIÓN ENTRE APRENDIZAJE AUTOMÁTICO Y PROFUNDO PARA LA PREDICCIÓN DE SERIES TEMPORALES

RODRÍGUEZ NUÑEZ MARTÍN¹, BALZARINI MONICA¹

¹ *Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba*

martinrnu@gmail.com

RESUMEN

La predicción precisa de series temporales es un desafío crucial en diversos campos como la economía, la meteorología y la salud pública, donde la precisión y la fiabilidad de las predicciones pueden tener un impacto significativo en la toma de decisiones y la calidad de vida de las personas. Este estudio busca comparar el rendimiento de modelos de aprendizaje automático y profundo en la predicción de series temporales. Se entrenaron dos modelos de aprendizaje profundo, uno basado en redes neuronales con memoria a largo y corto plazo (*LSTM*) y uno más sofisticado con arquitectura de codificador-decodificador (*encoder-decoder*) con *LSTM*, así como uno basado en el aprendizaje automático como lo es el árbol de gradiente aumentado (*XGBoost*). Para su entrenamiento y evaluación, se generaron 100.000 filas de datos simulados utilizando un modelo *ARIMA* introduciendo ruido aleatorio, creando series temporales. Los datos se dividieron en conjuntos de entrenamiento y testeo, normalizándose para garantizar el aprendizaje de los algoritmos. La evaluación del rendimiento se realizó mediante el cálculo de las métricas de *RMSE* y *MAE*. Los modelos de aprendizaje profundo, como los *LSTM*, están diseñados para capturar dependencias a largo plazo en los datos temporales, lo que les permite manejar secuencias con mayor precisión. En contraste, los modelos clásicos como *XGBoost*, aunque poderosos en muchas aplicaciones, no están inherentemente diseñados para capturar la naturaleza secuencial de los datos temporales. Sin embargo, son ampliamente utilizados para esta tarea dado al desconocimiento de sus limitaciones específicas en series temporales y gran versatilidad para diversas aplicaciones. Este trabajo busca contribuir a la comprensión de las ventajas de los modelos de aprendizaje profundo en la predicción de series temporales frente a los de aprendizaje automático, resaltando las limitaciones de estos últimos y destacando la importancia del algoritmo de aprendizaje en capacidad predictiva.

Palabras clave: ***aprendizaje profundo, series temporales, aprendizaje automático.***

CIENCIA DE DATOS PARA SABER SI LOS MEDICAMENTOS PUEDEN LLEGAR AL CEREBRO

DANTE N. STRATICO¹, MARIA LILIANA MIRANDA², VICTORIA RICHMOND^{2,3}, MARCELO OTERO¹

¹*IFIBA (CONICET-UBA) y Depto. de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

²*UMYMFOR (CONICET-Universidad de Buenos Aires)*

³*Depto. de Química Orgánica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

mlmiranda@qo.fcen.uba.ar

RESUMEN

La Barrera HematoEncefálica (BHE) es una estructura altamente selectiva que separa al sistema circulatorio del sistema nervioso central, protegiendo al cerebro de sustancias nocivas y patógenos, convirtiéndose a su vez en una limitante a la hora de encontrar nuevos medicamentos que logren acceder al cerebro. Esto tiene implicaciones no solo en las enfermedades neurodegenerativas, sino también en enfermedades psiquiátricas y en la administración de drogas con target en este órgano. Con el objetivo de obtener modelos de Machine Learning Supervisado capaces de predecir si un compuesto puede atravesar la BHE y entender mejor las diferencias, similitudes y propiedades de los compuestos que la atraviesan (BHE+) y los que no (BHE-), se utilizó la base de datos “Blood Brain Barrier Database” (B3DB) que cuenta con datos de permeabilidad de la BHE de alrededor de 6.000 compuestos. Esta base se caracterizó con 50 descriptores moleculares relacionados a propiedades fisicoquímicas de las moléculas. Posteriormente, se realizó la visualización conjunta de los dos grupos de compuestos, BHE- y BHE+, con diferentes tipos de técnicas de reducción de la dimensión: lineales como el Análisis de Componentes Principales (PCA) y no lineales como UMAP y t-SNE, para diferentes métricas y valores de sus hiperparámetros. Para lograr una mayor diferenciación en el espacio entre ambos grupos, se seleccionó un conjunto mínimo de descriptores: número de átomos de Oxígeno, Azufre y halógenos, proporción de oxígenos ácidos, masa molecular, área de Van der Waals, superficie polar relativa y aceptores de átomos de Hidrógeno, los cuales maximizan la separación en el espacio entre los BBB+ y los BBB- a través de su visualización con box-plots, violin-plots, histogramas y otros métodos de análisis exploratorios de datos. También se obtuvieron nuevas variables reducidas para entrenar los modelos de Machine Learning Supervisado mediante estas técnicas.

Palabras clave: **Barrera Hematoencefálica, Cs. de Datos, Descubrimiento de Drogas, PCA, t-SNE, UMAP.**



IMPACTO DEL TAMAÑO DE LA MUESTRA EN LA PRECISION DE MODELOS DE INTERPOLACION ESPACIAL. APLICACIONES EN AGRICULTURA

SUAREZ FRANCO¹, CÓRDOBA MARIANO^{1,2}, BALZARINI MÓNICA^{1,2}

¹ *Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba*

² *Grupo de Estadística Vinculado. Unidad de Fitopatología y Modelización Agrícola. (UFyMA-INTA-CONICET). Córdoba*

Suarezfranco@agro.unc.edu.ar

RESUMEN

Para ajustar modelos predicción robustos, es fundamental contar con un conjunto de datos de calidad. Sin embargo, el tamaño de la muestra suele estar limitado por los costos y tiempo de recolección, así como por los mecanismos de selección de los sitios de muestreo. En el caso de datos georreferenciados, el muestreo es más complejo dado la presencia de correlaciones espaciales y a la necesidad de representar la variabilidad espacial del proceso. En este trabajo se evalúa como el tamaño de la muestra afecta la precisión de la predicción espacial utilizando datos agrícolas en dos escalas espaciales: intralote y regional. A nivel intralote, se analizaron tres mapas de rendimiento de cultivos (soja, maíz y trigo) con 25,000 observaciones, comparando los métodos de kriging (KG) y bosques de regresión cuantílica (QRF). A nivel regional, se evaluó la probabilidad de presencia de enfermedades en patosistemas vegetales utilizando datos sobre la presencia/ausencia de patógenos y variables climáticas, comparando regresión logística (RL) y bosques aleatorios (RF) en contextos con y sin selección de covariables previo al ajuste de los modelos. Los tamaños de muestra se generaron utilizando el método Hipercubo Latino Condicionado. Los resultados mostraron que el rendimiento de los modelos mejora rápidamente con el tamaño de muestra, aunque la tasa de mejora disminuye con tamaños de muestra altos. A escala intralote, KG mostró un mejor desempeño con tamaños de muestra mayores, mientras que QRF demostró mayor robustez ante variaciones en el tamaño. A escala regional, RL presentó mejor desempeño con covariables predefinidas, pero RF superó a RL en precisión cuando no se realiza la selección de covariables previo al ajuste del modelo. La selección tanto del tamaño de la muestra y el modelo de predicción espacial son cruciales para optimizar la precisión de las predicciones espaciales.

Palabras clave: **modelos predictivos, aprendizaje automático, Hipercubo Latino Condicionado.**



REDES NEURONALES APLICADAS EN LA PREDICCIÓN DE LA DEMANDA QUÍMICA DE OXÍGENO QUE INGRESA A LA PLANTA DE TRATAMIENTO DE LÍQUIDO CLOACAL

VICTORIA MURILLO¹, MARIEL ZERBATO², JOSEFINA CARRIÓ³, LAURA MODINI², ANA PIZARRO², STELLA VAIRA³

¹*Tesinista de la Licenciatura en Saneamiento Ambiental, Cátedra Tratamiento de Efluentes, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral*

²*Cátedra Tratamiento de Efluentes, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral*

³*Departamento de Matemática y Laboratorio de Investigaciones y Servicios en Bioestadística (LISEB), Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral*

svaira@fbc.unl.edu.ar

RESUMEN

Evaluar los parámetros fisicoquímicos del agua residual urbana que ingresa a la planta de tratamiento de líquidos cloacales ayuda a anticipar problemas operativos y proporciona información crucial para la gestión del saneamiento y la protección del ambiente. El objetivo del trabajo fue construir un modelo estadístico que permita predecir la demanda química de oxígeno (DQO) del líquido cloacal que ingresa a la planta de tratamiento de la Ciudad de Santo Tomé, Santa Fe, considerando componentes físicos y químicos determinados en el agua residual urbana y el efecto de la temperatura ambiente. Se utilizaron técnicas estadísticas de la ciencia de datos con un doble rol: dejar un antecedente y estudio preliminar de técnicas estadísticas avanzadas, y aprovechar la dimensión del problema; $n=99$ y $p=13$. Teniendo en cuenta los parámetros químicos, físicos y la temperatura ambiental promedio, se realizó un análisis de componentes principales (ACP) y se observó que con cuatro componentes principales se podía explicar el 83,9% de la variabilidad total. Del análisis regresión de componentes principales (PCR) al predecir la DQO resultaron significativas las dos primeras componentes (65% de variancia explicada). Con Redes Neuronales, utilizando la misma base de datos y las mismas cuatro componentes principales que en PCR, fue posible obtener un modelo predictivo para DQO. Se puede concluir que ambos modelos resultan adecuados para predecir la DQO y que no hay grandes diferencias entre ellos en cuanto a su capacidad de predicción para este conjunto de datos, ya que, por ejemplo, para un valor medido de DQO de $250\text{mgO}_2/\text{L}$ el modelo de redes neuronales predijo un valor de $241\text{mgO}_2/\text{L}$ y en PCR de $234\text{mgO}_2/\text{L}$. Es importante remarcar que fueron entrenados con la totalidad de los datos. Se recomienda aplicar una técnica de imputación de datos y volver a construir los modelos con la posibilidad de dividir la base en muestras para el entrenamiento y para el test, comparar con métricas con respecto al modelo no imputado. Se procesaron los datos con RStudio y la librería neuralnet.

Palabras clave: **agua residual urbana, modelos estadísticos, análisis multivariado.**

Ciudad de Buenos Aires, octubre 2024



Bases de Datos y Datos Masivos



ESTIMACIÓN DE CORRELACIONES ENTRE RENDIMIENTOS DE CULTIVOS PROVENIENTES DE REGISTROS EMPRESARIALES

LUIS FRANK¹, FACUNDO LANDAJÓ², CRISTINA RAS²

¹*Departamento de Métodos Cuantitativos y Sistemas de Información, Facultad de Agronomía,
Universidad de Buenos Aires*

²*Departamento de Economía, Desarrollo y Planeamiento Agrícola, Facultad de Agronomía,
Universidad de Buenos Aires*

frank@agro.uba.ar

RESUMEN

Se propone una metodología de estimación de coeficientes de correlación entre rendimientos de cultivos a nivel de empresa. El método se basa en la combinación de una regresión lineal y su versión revertida en un único modelo, uno de cuyos parámetros es el coeficiente de interés. El método resulta particularmente útil para estimar correlaciones en situaciones donde los tamaños muestrales son extremadamente pequeños, como ocurre cuando los planes de producción de la empresa cambian año tras año, y ciertos cultivos rara vez se producen juntos. A lo largo del artículo se realiza un ejercicio de estimación con datos reales, y los resultados se comparan con estimaciones previas calculadas a partir de estadísticas públicas. Se concluye que las correlaciones calculadas a nivel de empresa son, en general, mayores que las calculadas a nivel agregado, al menos en el área de estudio. Asimismo, se halla evidencia de que las correlaciones entre rendimientos no son fijas para el mismo sitio o empresa, sino que pueden variar en función del ciclo climático ENSO, al menos para ciertos pares de cultivos. Sobre la base de estos resultados, se proponen algunas guías para construir matrices de correlaciones, cuando no se disponga de datos a nivel de empresa.

Palabras clave: *simulación, correlación, agricultura.*



ESTUDIO DE ASOCIACIÓN DE GENOMA COMPLETO PARA MIOCARDIOPATÍA CHAGÁSICA EN LA PROVINCIA DE CÓRDOBA: RESULTADOS PRELIMINARES

MARTÍN BERNAUS¹, SEBASTIÁN FILIPIGH¹, FRANCO SUÁREZ¹, DANIELA VELÁZQUEZ LÓPEZ², ROMINA BLASCO², SILVINA LO PRESTI², CECILIA BRUNO¹

¹*Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba. Grupo vinculado de Estadística y Biometría Unidad de Fitopatología y Modelización Agrícola (UFyMA-INTA-CONICET), Córdoba*

²*Centro de Estudio e Investigación de la Enfermedad de Chagas y Leishmaniasis, FCM, INICSA-CONICET-UNC, Córdoba.*

martin.bernaus@mi.unc.edu.ar

RESUMEN

La cardiomiopatía chagásica crónica (CCC) es una de las manifestaciones más severas de la enfermedad de Chagas y su desarrollo está influenciado por factores inmunogenéticos del paciente. El objetivo del trabajo es realizar un estudio de asociación de genoma completo (GWAS) sobre pacientes chagásicos cordobeses, evaluando el desempeño de modelos estadísticos tradicionales y avanzados para la identificación de regiones genéticas asociadas al desarrollo de la CCC. Se trabajó con una base de datos de 4,5 M de SNPs extraídos de los datos genéticos de casos y controles para la CCC. Se realizó un análisis de componentes principales (ACP) con test de Tracy-Widom (ACP-TW) para estimar estructura genética poblacional (EGP) y se aplicaron ocho modelos de asociación estadística, incluyendo la regresión logística, modelos lineales generales (GLM) y modelos lineales mixtos (MLM), GLMM, MLMM, SUPER, FarmCPU y Blink. Estos últimos utilizan métodos de máxima similitud y BIC para lograr un alto poder estadístico y eficiencia computacional. El ACP evidenció la presencia de EGP y fue incorporado en los modelos. Se identificaron dos regiones genéticas de interés, ubicadas en los cromosomas 3q12.2 y 14q24.3, que fueron señaladas como las asociaciones más significativas en todas las configuraciones. La regresión logística, FarmCPU y Blink fueron los más exitosos para identificar 11 SNPs entre ambas regiones como putativamente asociados ($p < 1.10^{-6}$) o fuertemente asociados ($p < 5.10^{-8}$) a la CCC. Su asociación representa un hallazgo novedoso: estas regiones no han sido previamente vinculadas a la enfermedad, lo cual implica la necesidad de profundizar el estudio sobre su ubicación y función fisiológica, y de considerar un posible efecto de las comorbilidades. Por otra parte, se espera que la aplicación de modelos más complejos, basados en aprendizaje automático, supere la capacidad de detección de asociaciones significativas de los algoritmos estadísticos utilizados hasta el momento.

Palabras clave: **GWAS, SNPs, modelos mixtos.**



CORRELACIONES ENTRE RENDIMIENTOS DE CULTIVOS PAMPEANOS ESTIMADAS A PARTIR DE ESTADÍSTICAS PÚBLICAS

LUIS FRANK¹

¹*Departamento de Métodos Cuantitativos y Sistemas de Información, Facultad de Agronomía,
Universidad de Buenos Aires*

frank@agro.uba.ar

RESUMEN

La simulación del resultado económico de una empresa agrícola requiere la introducción de una matriz de correlaciones entre rendimientos de cultivos. No existe, sin embargo, una metodología comúnmente aceptada por los analistas de riesgo agropecuario para estimar dicha matriz, ni una matriz teórica de referencia que la sustituya en caso de que no pueda ser estimada por falta de datos. En el trabajo se presenta una guía metodológica para la estimación de matrices a nivel de partido o departamento, y se estima una matriz de consenso para toda la región pampeana a partir de rendimientos registrados por la ex SAGPyA (con discontinuidades) durante 54 campañas agrícolas en 10 distritos de la región. Los resultados ubican a las correlaciones entre soja, maíz y sorgo en el orden de 0,5; entre girasol y los demás cultivos de verano, entre 0,2 y 0,3; y entre trigo y los cultivos de verano, entre 0,1 y 0,2. La correlación entre soja de primera y soja de segunda se ubica en 0,7. Estas estimaciones constituyen, en principio, cotas inferiores para la construcción de matrices a nivel subregional o local, de acuerdo a la hipótesis que sugiere que a mayor homogeneidad productiva del área bajo estudio, mayor correlación entre rendimientos. Sin embargo, la comparación de la matriz de consenso con matrices empíricas de diverso origen (registros empresariales o modelos de simulación agronómica) no validan esta afirmación. Al final del trabajo se discuten posibles causas de esta contradicción.

Palabras clave: *simulación, correlación, agricultura.*



Datos Espaciales

MODELANDO LA RESPUESTA DEL CULTIVO A LA APLICACIÓN DE INSUMOS POR ZONAS DE MANEJO EN ENSAYOS A ESCALA DE LOTE

CARLOS AGUSTIN ALESSO¹, PATRICIA MELINA ACETTA², MAXIMILIANO SABINO³, GONZALO PARRA³, PABLO ARIEL CIPRIOTTI⁴

¹*CiAgro Litoral, CONICET, Facultad de Ciencias Agrarias, Universidad Nacional del Litoral*

²*Facultad de Ciencias Agrarias, Universidad Nacional del Litoral*

³*Departamento de Desarrollo, Grupo Don Mario (GDM)*

⁴*Facultad de Agronomía – IFEVA, Universidad de Buenos Aires / CONICET*

calesso@fca.unl.edu.ar

RESUMEN

La agricultura de precisión (AP) asume la interacción entre la respuesta del cultivo a las prácticas de manejo y la heterogeneidad de ambientes o zonas intra-lote. Los ensayos a escala de lote pueden ayudar a verificar dicha hipótesis y establecer recomendaciones de manejo sitio-específicas. Las características de los datos generados en estos ensayos requieren técnicas estadísticas avanzadas para lidiar con estructuras de correlación espacial dentro de las unidades experimentales. Nuestro objetivo fue explorar alternativas de análisis de ensayos a escala de lote conducidos con tecnologías de AP. Se analizaron 10 ensayos de respuesta del cultivo de maíz a la densidad de siembra (DS) y fertilización nitrogenada (N) en el sur de Córdoba y oeste de Buenos Aires. Todos los lotes se sembraron con el mismo híbrido y contaban con zonas de manejo (ZM) de alta y baja productividad. Los tratamientos se dispusieron de manera sistemática y sin repetición en sentido perpendicular al mayor gradiente de variación. Se ajustaron 8 modelos mixtos para cada ensayo considerando ZM, N, modelo cuadrático de DS como factores cruzados fijos, y distintas estructuras de efectos aleatorios para modelar la correlación espacial con distinto nivel de detalle, desde sólo considerar la parcela del ensayo hasta tener en cuenta las coordenadas planares del registro del monitor de rendimiento a distintos niveles (parcela o ensayo). En 8 de los 10 ensayos la inclusión de la correlación espacial, en función del orden en la parcela o de las coordenadas a nivel de ensayo o por parcela, removió el efecto de la respuesta a insumos por ZM que fuera observada en los modelos no espaciales. Sólo en un caso la recomendación de DS varió marginalmente según N y ZM. Claramente, la especificación incorrecta de los efectos aleatorios omitiendo la estructura de correlación espacial tiene alto impacto en las recomendaciones.

Palabras clave: ***agricultura por ambientes, correlación espacial, respuesta sitio-específica, geoestadística, modelos mixtos***



ANÁLISIS DE PATRONES DE PUNTOS PARA DETECTAR ASOCIACIONES ESPACIALES ENTRE HORMIGAS DE DESIERTO Y CARACTERÍSTICAS DE SU ENTORNO

ANA L. CAO¹, FERNANDO A. MILESI², RODRIGO G. POL¹³, LUCÍA VULLO¹, M. FLORENCIA MIRETTI¹, JAVIER LOPEZ DE CASENAVE¹

¹ *Ecodes, Facultad de Ciencias Exactas y Naturales e IEGEBA-CONICET, Universidad de Buenos Aires*

² *Instituto de Investigaciones en Biodiversidad y Medioambiente (INIBIOMA-CONICET) - Universidad Nacional del Comahue*

Instituto Argentino de Investigaciones de las Zonas Áridas (IADIZA-CONICET), Universidad Nacional de Cuyo

annylau2612@gmail.com

RESUMEN

La estadística espacial exploratoria y el ajuste de parámetros de modelos estocásticos a patrones de puntos permiten detectar y sopesar la importancia del componente espacial de diferentes procesos ecológicos. Buscamos y determinamos la ubicación espacial de colonias de hormigas de los géneros *Pheidole* y *Pogonomyrmex* en parcelas (2D) y en transectas (1D) en varias temporadas en la porción central del desierto del Monte (Reserva de la Biosfera de Ñacuñán, Mendoza). Estudiamos las propiedades de primer (intensidad) y segundo orden (autocorrelación espacial) de los procesos espaciales que podrían haber generado los patrones observados, poniendo a prueba los supuestos más parsimoniosos: homogeneidad de la intensidad (pruebas Kolmogorov-Smirnov de bondad de ajuste, AIC de modelos Poisson con y sin coordenadas como predictores) y ausencia de autocorrelación espacial (funciones de Ripley y de correlación de pares con simulaciones Monte Carlo). En función de esos resultados exploramos la relación de la ubicación de las colonias con variables ambientales potencialmente claves. Con análisis 1D (transectas como redes lineales de un segmento) concluimos que el agregamiento intra-específico y la segregación interespecífica de las tres especies de *Pogonomyrmex* simpátricas se debe a sus relaciones contrastantes con las variaciones de la permeabilidad del suelo y la topografía (modelos Poisson de intensidad heterogénea con esas variables predictoras), que persisten entre años, y no con interacciones positivas/negativas entre sus colonias. Con análisis 2D (comparación de modelos Poisson y modelos Gibbs con interacción positiva/negativa entre puntos) explicamos la disposición de colonias de *Pheidole bergi* en función de la distancia a los bordes entre parches con y sin vegetación leñosa y a la intensidad de las interacciones entre colonias, que cambiaron entre años de alta y baja densidad. Con estas técnicas logramos tanto detectar como descartar factores bióticos y abióticos relevantes para la nidificación y la supervivencia de estas hormigas de desierto.

Palabras clave: ***estadística espacial, spatstat, patrones de puntos, hormigas***

Ciudad de Buenos Aires, octubre 2024

ACUICULTURA 4.0: ESTIMACIÓN BIODINÁMICA DE LA PRODUCCIÓN DE MEJILLONES

VERÓNICA COY¹, PABLO TURJANSKI^{1,2}, HENRIQUE KNACK DE ALMEIDA^{3,4}

¹*Departamento de Computación, Fac. de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

²*Instituto de Ciencias de la Computación (ICC). CONICET-Universidad de Buenos Aires.*

³*Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

⁴*Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN). CONICET-Universidad de Buenos Aires*

vcoy@dc.uba.ar

RESUMEN

La acuicultura constituye globalmente el sector de mayor crecimiento en producción de alimentos. Sin embargo, carece de un conjunto actualizado de principios que garanticen su apropiada expansión e intensificación. En ese sentido, desde el punto de vista socioeconómico y medioambiental se hace imprescindible la actualización de procesos productivos mediante la incorporación de nuevas tecnologías. Inspirados en la idea de la "Industria 4.0", estamos desarrollando "Acuicultura 4.0", una plataforma digital para el sector acuícola, en la cual se adoptan herramientas tecnológicas para optimizar la gestión, monitoreo en tiempo real y productividad de los cultivos. Con base en dispositivos de sensores ambientales, desarrollados de manera ad hoc y conectados en red, se registrarán las variaciones en la temperatura, y en las concentraciones de oxígeno disuelto y clorofila a en relación a la proximidad de la costa y a la profundidad de distintos puntos del cultivo. A partir de los datos ambientales se investigarán sus efectos sobre las respuestas metabólicas y fisiológicas relacionadas con la productividad de *Mytilus edulis platensis*, usando como caso testigo un cultivo de líneas flotantes en Bahía Camarones. Simultáneamente, los resultados se emplearán en un modelo bioenergético para la estimación del rendimiento del cultivo. En síntesis, el presente trabajo propone el desarrollo de una plataforma digital que permitirá el monitoreo ambiental, así como la visualización del análisis productivo (poniendo a disposición herramientas de estadística descriptiva que posibiliten al usuario el entrecruzamiento de múltiples variables) y la estimación del crecimiento basada en la ecuación de von Bertalanffy Growth Function (vBGF) en un sistema de cultivo de mejillones. Dicha plataforma ofrecerá una herramienta sencilla y robusta para la optimización de la gestión de cultivos de productos marinos y permitirá evaluar y determinar nuevas áreas marítimas del litoral argentino aptas para el cultivo sustentable de bivalvos con valor comercial.

Palabras clave: **acuicultura 4.0; estimación de la producción; sensores; plataforma digital; bivalvos marinos; von Bertalanffy Growth Function.**

Ciudad de Buenos Aires, octubre 2024

DETECCIÓN DE CAMBIOS EN LA DISTRIBUCIÓN ESPACIAL DE UNA COMUNIDAD MARINA MEDIANTE MODELOS NULOS

CUESTA NÚÑEZ JOSEFINA^{1,2,3}, OCAMPO REINALDO MATIAS^{2,3,4}, ROMERO MARIA ALEJANDRA^{2,3,4}, FLOREZ LAURA^{2,3}, GONZALEZ RAUL^{2,3,4}, SVENDSEN GUILLERMO^{2,3,4}.

¹*Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

²*Centro de Investigación Aplicada y Transferencia Tecnológica en Recursos Marinos "Almirante Storni" (CIMAS)*

³*CONICET (CONICET)*

⁴*Escuela Superior de Ciencias Marinas, Universidad Nacional del Comahue*

josefinacuestan@gmail.com

RESUMEN

En el Golfo San Matías se realizan campañas de investigación pesquera utilizando un diseño regular de 41 lances de pesca de 30 minutos. En 2022, la duración de los lances se redujo a 15 minutos, lo que podría afectar la detectabilidad de ciertas especies. Este estudio evalúa si la distribución de las especies en 2022 difiere de la observada entre 2006 y 2018, considerando esta reducción del esfuerzo de muestreo. Usamos un modelo nulo para identificar especies presentes en un número de lances diferente al esperado. Valores atípicos positivos indicarían una expansión de la especie dentro del golfo, mientras que valores negativos podrían sugerir una retracción o menor detectabilidad debido al menor esfuerzo de muestreo. Para las especies con valores atípicos negativos, verificamos si los resultados de 2022 se correspondían con tendencias negativas de frecuencia de ocurrencia espacial (FO) del período previo (2006 - 2018) mediante regresiones lineales. Para construir el modelo nulo, calculamos la FO esperada de cada especie haciendo 1000 remuestreos aleatorios de 41 lances para el período 2006-2018. Luego comparamos los valores observados en 2022 con las FO esperadas. Identificamos valores atípicos usando el rango intercuartílico (IQR), considerando como atípicas aquellas FO por debajo de $Q1 - 1,5 * IQR$ o por encima de $Q3 + 1,5 * IQR$ de las distribuciones esperadas. Encontramos tres especies con una FO mayor a la esperada, sugiriendo expansión dentro del golfo. Diez especies mostraron una FO menor a la esperada: cinco con tendencias previas decrecientes (indicando que la retracción espacial continuó en 2022) y cinco con tendencias previas positivas (posiblemente afectadas por la reducción del esfuerzo de muestreo o cambios en su distribución). Esta metodología nos permitió detectar cambios espaciales en la comunidad demersal a pesar del menor esfuerzo de muestreo.

Palabras clave: ***ecología marina; frecuencia de ocurrencia; esfuerzo de muestreo***

DESIGUALDADES TERRITORIALES EN LA REALIZACIÓN DE MAMOGRAFÍAS: ANÁLISIS A PARTIR DE MODELOS BAYESIANOS

MAGDALENA DAMIANI QUIROZ¹, RAÚL MEJÍA^{2,3}, GERARDO CUETO^{1,4}, MARÍA SOLEDAD FERNÁNDEZ^{1,5}

¹Grupo de Bioestadística Aplicada, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

²Centro de Estudios de Estado y Sociedad (CEDES)

³Hospital de Clínicas, Universidad de Buenos Aires

⁴Instituto de Ecología, Genética y Evolución (IEGEB), Universidad de Buenos Aires-CONICET

⁵Instituto de Cálculo, Universidad de Buenos Aires -CONICET

maguidamiani@gmail.com

RESUMEN

El cáncer de mama representa el 32% del total de incidencia de cáncer en Argentina, su detección temprana aumenta la supervivencia. Con el objetivo de identificar desigualdades territoriales relacionadas con la realización de mamografías en mujeres (edades 50-69) en Argentina, se trabajó con datos de la Encuesta Nacional de Factores de Riesgo (2005-2009-2013-2018). Se generaron mapas de realización de mamografía en Argentina a partir del ajuste de modelos espaciales bayesianos para dos variables respuesta a escala provincial para cada edición: proporción de mujeres que reportaron haberse realizado una mamografía alguna vez respecto al total de mujeres encuestadas (“acceso”), y proporción de mujeres que reportaron haberse realizado una mamografía en los últimos dos años respecto al total de mujeres encuestadas que se realizaron una mamografía alguna vez (“seguimiento”). Se incorporó a la provincia como factor aleatorio y la edad promedio de las mujeres (centrada) como variable control. Se asumió una distribución normal para los coeficientes y una distribución gamma para las varianzas. Se evidenciaron desigualdades territoriales que se atenuaron con el paso del tiempo, donde las provincias del norte presentan menor proporción de mujeres que alguna vez realizaron una mamografía (2005: 35%-65%, 2018: 65%-85%) y de mujeres que se realizaron una en los últimos dos años (2005: 60%-75%, 2018: 65%-75%) respecto al resto de las provincias (CABA, 2005 y 2018, acceso: 85%-95%, seguimiento: 80%-85%). El rango de variación de valores predichos entre provincias fue mayor para la variable de acceso (35%-95%) respecto a la de seguimiento (60%-85%), al igual que el porcentaje de variabilidad total explicada por el componente espacial (acceso: 61%-75%, seguimiento: 25%-60%). Estos resultados evidencian la existencia de desigualdades territoriales, principalmente en el acceso y luego en el seguimiento, en la realización de mamografías entre las provincias de Argentina.

Palabras clave: **mamografía, modelos espaciales bayesianos, INLA, análisis espacio-temporal**

Ciudad de Buenos Aires, octubre 2024

ESTUDIO DE LA VARIABILIDAD TEMPORAL DE PASTIZALES INVADIDOS USANDO SERIES TEMPORALES DE IMÁGENES SENTINEL-2

DIAZ VILLA MVE¹², CAGNONE F¹², PARONETTO J¹², CRISTIANO PM¹², BRUZZONE OA³

¹ *Laboratorio de Ecología Funcional, Instituto de Ecología, Genética y Evolución (IEGEB), Universidad de Buenos Aires-CONICET*

² *Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

³ *Instituto de Investigaciones Forestales y Agropecuarias Bariloche, INTA y CONICET, Bariloche, Río Negro*
virginiadv@ege.fcen.uba.ar

RESUMEN

Los pastizales de la provincia de Buenos Aires se encuentran fuertemente amenazadas por invasiones de especies leñosas. El objetivo de este trabajo fue realizar una clasificación funcional a partir de la variabilidad temporal intra- e inter-anual de la vegetación en un pastizal invadido por Acacia negra. Para esto, se analizaron series temporales del índice verde (NDVI) estimado a partir de imágenes Sentinel-2. Utilizando la plataforma Google Earth Engine se obtuvieron un total de 405 imágenes entre mayo 2018 y julio 2024, abarcando un área de 307 ha en el Parque Nacional Ciervo de los Pantanos. Todos los análisis se realizaron a nivel de pixel. En primer lugar, se utilizó un modelo de mezcla gaussiana para eliminar observaciones espurias de la serie temporal. Luego, se realizó un análisis espectral usando periodogramas de Lomb-Scargle para caracterizar la distribución de la variabilidad entre distintas longitudes de onda (desde 1 mes hasta 5 años). Los análisis espectrales se compusieron en una matriz de periodogramas que, a su vez, fue utilizada para realizar análisis arquetípicos para definir diferentes tipos de tipologías o arquetipos. La clasificación resultó en la identificación de 5 tipologías: cuatro de ellas presentaron una fuerte componente anual, con mayor o menor aporte de bajas frecuencias (menores a 1/año), mientras que uno de ellos presentó un mayor componente bianual (2/año), aunque baja variabilidad explicada o potencia. El análisis de arquetipos permitió identificar distintas tipologías y su distribución en el área, así como caracterizar su comportamiento. La distribución de las tipologías parece representar una compleja interacción entre la cobertura vegetal y factores como el microrelieve y la frecuencia de incendios.

Palabras clave: **análisis de arquetipos, periodograma de Lomb-Scargle, NDVI**



INTERPOLACIÓN ESPACIAL DE MATERIA ORGÁNICA DEL SUELO A ESCALA DE LOTE USANDO QUANTILE REGRESSION FOREST

GARCÍA SELEME F.A.¹², CÓRDOBA M.A.¹², PACCIORETTI P.¹³⁴, BALZARINI M.¹²

¹*Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba*

²*Grupo de Estadística Vinculado. Unidad de Fitopatología y Modelización Agrícola. (UFyMA-INTA-CONICET). Córdoba*

³*Instituto Mario Gulich. Comisión Nacional de Actividades Espaciales. Universidad Nacional de Córdoba*

⁴*University of Nebraska. Department of Agronomy and Horticulture. Lincoln, Estados Unidos*

fabrizio.garcia.seleme@mi.unc.edu.ar

RESUMEN

La materia orgánica del suelo (MO) es un componente esencial para la fertilidad y productividad agrícola, por lo que estimar su distribución espacial con precisión es esencial para un manejo eficiente y sostenible del suelo. Tradicionalmente, la interpolación espacial se ha realizado mediante métodos geoestadísticos como Kriging, que requieren cumplir supuestos distribucionales y contar con una cantidad suficiente de datos para ajustar semivariogramas confiables. Los algoritmos de aprendizaje automático, como Quantile Regression Forest (QRF), ofrecen una alternativa prometedora al ser más flexibles respecto a estos supuestos y facilitar la integración de datos provenientes de múltiples lotes agrícolas. Sin embargo, QRF no considera la estructura de autocorrelación espacial en el ajuste del modelo. Para abordar esta limitación, han surgido nuevas alternativas como Quantile Regression Forest Spatial Interpolation (QRFI), que incorpora la información espacial utilizando como covariables los valores de la variable respuesta observada en los sitios más cercanos a los puntos de predicción. En este trabajo, se comparó QRFI con Kriging en su capacidad para mapear la distribución intralote de MO. Se utilizaron 163 puntos georreferenciados de MO, muestreados en cinco lotes agrícolas. Los modelos se ajustaron utilizando como variables explicativas la conductividad eléctrica aparente del suelo, elevación y rendimiento del cultivo, procesadas en una grilla común de 10 × 10 m. El error de predicción se estimó mediante validación cruzada k-fold (k=10). Los resultados mostraron que QRFI superó a Kriging en términos de error de predicción (RMSE). QRFI se representa como una herramienta prometedora para el mapeo digital del suelo a escala fina.

Palabras clave: aprendizaje automático, autocorrelación espacial, mapeo digital del suelo.

COMPARACIÓN DE MÉTODOS DE VALIDACIÓN CRUZADA EN SERIES DE TIEMPO SIMULADAS UTILIZANDO MODELO ARIMA PARA PRONÓSTICO DE PRECIO

HAIQUE, A.K.¹ RUBIO, N.² MALASPINA, M.³ GIMENEZ, G.⁴

^{1 2 4} *Departamento de Estadística, Facultad de Economía y Administración, Universidad Nacional del Comahue, Neuquén*

³ *Departamento de Economía, Facultad de Economía y Administración, Universidad Nacional del Comahue, Neuquén*

anahaique@gmail.com

RESUMEN

El objetivo de este estudio fue comparar dos enfoques de validación cruzada en series temporales transformadas y simuladas, utilizando un modelo ARIMA. La serie original utilizada fue el Precio Promedio del Cordero publicada por el INTA (PPCord), que muestra un comportamiento exponencial y fue transformada para linealizarla. Se generaron 100 series simuladas a partir de la serie original añadiendo ruido aleatorio mediante una distribución normal, manteniendo su estructura. Se evaluaron dos métodos de validación cruzada (VC): (1) origen fijo para el grupo de entrenamiento y grupo de testeo deslizante, y (2) ventanas deslizantes tanto en el grupo de entrenamiento como en el de testeo. En el primero, el grupo de entrenamiento crece progresivamente con cada partición, mientras que en el segundo, se mantuvo un tamaño de entrenamiento y testeo constantes que se desliza en el tiempo. Para evaluar el rendimiento del modelo ARIMA, en cada uno de los métodos de validación, se utilizaron las métricas: Error Absoluto Medio, Error Porcentual Absoluto Medio y la Raíz del Error Cuadrado Medio, que miden la precisión de los pronósticos en el grupo de testeo. Finalmente, se compararon los resultados obtenidos por ambos métodos de validación cruzada en función de las métricas de error mencionadas y el tiempo de renderización, con el objetivo de determinar cuál de los dos enfoques es más efectivo para trabajar con series temporales similares al PPCord. Se concluyó que los resultados son similares pero, el tiempo de renderización promedio del segundo es 17% menor lo que resulta beneficioso en entornos con grandes volúmenes de datos. Por otra parte, el método de origen fijo permite aprovechar al máximo la información disponible, ya que el conjunto de entrenamiento crece con cada partición, lo que puede resultar en modelos más robustos en ciertos contextos.

Palabras clave: ***método de origen fijo, método de ventana deslizante, simulación de series de tiempo.***

ANÁLISIS DE LA VARIABILIDAD ESPACIO-TEMPORAL DEL RENDIMIENTO EN UN SITIO CON VID VAR. CABERNET SAUVIGNON PARA PRONÓSTICOS DE PRODUCCIÓN Y ZONAS CON CALIDADES ENOLÓGICAS DIFERENCIADAS

LLERA, J.¹; MARTINENGO, N.¹; MORELLI, CI.¹; MARTÍNEZ, E.¹; MAURE, E.¹; GALIOTTI, H.²; CORVALAN, A.³; NAZRALA, J.²; SOSA, M.²; DRAQUE, A.²

¹ Cátedra de Cálculo Estadístico y Biom., Facultad de Ciencias Agrarias, Universidad Nacional de Cuyo

² Cátedra de Enología, Facultad de Ciencias Agrarias, Universidad Nacional de Cuyo

³ Cátedra de Viticultura, Facultad de Ciencias Agrarias, Universidad Nacional de Cuyo

jllera@fca.uncu.edu.ar

RESUMEN

Los mapas de sitio de rendimientos y covariables generados con un muestreo sistemático, determinan la variabilidad espacio-temporal de una parcela de vid Cabernet Sauvignon. Esto permite diseñar estrategias de muestreos probabilísticos para pronósticos parcelarios de producción potencial. Con la determinación de la variación espacial del sitio, se desean obtener pronósticos de mayor precisión y exactitud en sus estimadores. Además los mapas podrían identificar criterios de cosechas diferenciadas para mejores calidades enológicas. En un marco de 5155 plantas, se estructura una malla de alta densidad de puntos (2.2%) durante 2022-2024. En cada punto de muestreo, se evaluaron las cargas de órganos productivos y, mediante kriging ordinario, se generan mapas con la variación espacial. Se estratifican a las plantas según su cargas de racimos en categorías: buena, regular y poca. Desde la malla se aplican dos estrategias de muestreo aleatorios para estimar el peso individual de los racimos en los pronósticos antes de cosechar: una con aleatorio simple y otra estratificando. En pronósticos tempranos se emplearon modelos de razón y con información de campañas anteriores. Para una microvinificación diferenciada, se muestrearon y cosecharon plantas por categorías. Los mapas indicaron sectores diferentes de producción en el sentido Oeste-Este ($p < 0.001$). La variación temporal, indicó una inestabilidad en el sector SE. Para el periodo, los pronósticos antes de cosecha con la primera estrategia resultaron con errores de estimación y de pronósticos menores al 10% prefijado, al igual que los tempranos en 2023. La microvinificación diferenciada resultó significativo en su contenido de polifenoles totales ($p = 0.0064$). Los pronósticos antes de cosecha con muestreos aleatorios simples sobre la malla, resultaron con errores similares o menores que al estratificar. Los mapas de sitios con la carga de racimos y la determinación de su variación espacio-temporal, identifican tempranamente sectores diferenciados de producción potencial. La cosecha diferenciada por sectores permitiría diferenciar calidades enológicas.

Palabras clave: ***Pronóstico de producción-Muestreo-Manejo de sitio específico***

Ciudad de Buenos Aires, octubre 2024

ANÁLISIS DE LAS VARIACIONES TEMPORALES EN LAS BRECHAS DE NIVEL SOCIOECONÓMICO Y GÉNERO EN EL CONSUMO DE FRUTAS Y VERDURAS EN LA POBLACIÓN ADULTA DE ARGENTINA A PARTIR DE DATOS PÚBLICOS

MATHIAS JOEL LOPEZ¹, BRUVERIS ANA PAULA², PEREZ ADRIANA¹, FERNANDEZ MARIA SOLEDAD^{1,3}

¹Grupo de Bioestadística Aplicada, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

² Facultad de Ciencia y Tecnología de los Alimentos, Universidad Nacional del Comahue

³ Instituto de Cálculo, (Universidad de Buenos Aires-CONICET)

mathias.j.lopez@gmail.com

RESUMEN

El bajo consumo de frutas y verduras constituye un factor de riesgo para las enfermedades crónicas no transmisibles. Se ha observado un menor consumo en grupos socioeconómicamente desfavorecidos, y para grupos de mayor Nivel socioeconómico (NSE), un mayor consumo en mujeres respecto a hombres. Para analizar variaciones temporales en estas brechas, se trabajó con datos de la Encuesta Nacional de Factores de Riesgo (2009-2013-2018, muestra de 94.467 adultos de centros urbanos de Argentina) y de Indicadores Básicos provinciales. Se modeló el cumplimiento (cumple/no_cumple, Bernoulli) de consumo mínimo de frutas y verduras diario recomendado ("CFyVm", correspondiente 400 gramos, OMS) ajustando Modelos Lineales Generalizados y Mixtos. Las variables predictoras fueron el año de realización de la encuesta (factor), y, a nivel individual, el nivel de instrucción alcanzado (primario-secundario-universitario/terciario), género y rango de edad; a nivel hogar, el quintil de ingresos y la presencia/ausencia de necesidades básicas insatisfechas (NBI); y a nivel provincial el %hogares con NBI (terciles). Se ajustaron modelos alternativos con las predictoras mencionadas y diferentes interacciones que permitan analizar si las brechas previamente reportadas se modificaron en el tiempo. Se seleccionó el modelo que incluyó a las interacciones $\text{genero} * \text{quintil_ingresos}$ y $\text{quintil_ingresos} * \text{año}$, y el análisis de esta última mostró que la brecha en el CFyVm, según quintil de ingresos, se incrementó en el tiempo (OR $Q5/Q1_{2009}=1.4$ [IC_{95%}=1.3;1.5]; OR $Q5/Q1_{2013}=1.6$ [IC_{95%}=1.5;1.8]; OR $Q5/Q1_{2018}=2.0$ [IC_{95%}=1.8;2.2]). El resto de las predictoras se relacionaron con el CFyVm de la manera esperada, aumentando el CFyVm en grupos de mayor NSE y en mujeres de mayor NSE. Estos resultados aportan evidencia sobre el incremento en el tiempo de la brecha en el CFyV entre NSE en población adulta de Argentina.

Palabras clave: **alimentación saludable - inequidades - análisis temporal – brecha - modelos lineales generalizados y mixtos.**



DIVERSIDAD- β Y REGISTRO FÓSIL, MITIGACIÓN DE PSEUDOPATRONES

RODRIGO S. MARTÍN¹, NATALIA GARCÍA CHAPORI¹ & CECILIA LAPRIDA¹

¹Laboratorio de Sondeos de Ambientes Continentales y Marinos (SACMa), IDEAN, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires-CONICET

rodrigomartin88@gmail.com

RESUMEN

Uno de los principales inconvenientes asociados al análisis de la diversidad- β en asociaciones fósiles es la presencia de pseudopatrones, registros de extinciones/colonizaciones locales que no reflejan dinámicas ecológicas, sino que resultan de procesos aleatorios, estrategias inadecuadas de muestreo, falta de resolución taxonómica y/o la preservación diferencial de las especies. Debido a su propia naturaleza, muchas veces no es posible mejorar el muestreo por la irrepetibilidad o falta de disponibilidad de muestras. Dos de las estrategias más utilizadas para evitar pseudopatrones es la eliminación del análisis de las especies raras y el cálculo de medias móviles. Aquí nos proponemos analizar la validez de estas estrategias. Para ello se diseñó un experimento computacional en el que se realizaron 30 muestreos de 11 niveles estratigráficos en una sucesión que presenta un aumento exponencial en la diversidad. Se calculó la diversidad a partir de las matrices obtenidas mediante cuatro métodos. Dos de ellos trabajan sin sesgo: la extracción de especies raras y el cálculo de la media móvil. Los otros dos métodos consideran algunos valores como dudosos (p.ej., falsos positivos) y los extraen, resultando en matrices con celdas vacías para luego eliminar los niveles sin datos o bien sustituirlos por el promedio de los dos niveles suprayacentes y los dos infrayacentes. Para cada modelo se calculó el índice de Sørensen y se comparó la variación temporal del promedio de la diversidad contra su evolución teórica. Los valores obtenidos con eliminación de especies raras y sustitución por promedios son los que más se aproximaron a los valores medios teóricos de la diversidad, pero solo la sustitución reflejó tendencias semejantes.

Palabras clave: **Diversidad- β ; Pseudopatrones, Paleoecología.**



MODELOS GEOESTADÍSTICOS Y DE SERIES DE TIEMPO PARA ANALIZAR LA CALIDAD DEL AIRE EN LA CIUDAD DE SAN SALVADOR DE JUJUY

MARCELA L. NATAL¹, DANIELA A. MOLINARI^{1,2}, MARCOS VAIRA³, MARCOS A. E. CHAPARRO⁴

¹Centro Marplatense de Investigaciones Matemáticas (CEMIM), CICPBA, Universidad Nacional de Mar del Plata

²CONICET

³Instituto de Ecorregiones Andinas (INECOA-CONICET), Universidad Nacional de Jujuy

⁴Centro de Investigaciones en Física e Ingeniería del Centro de la Provincia de Buenos Aires (CIFICEN), Universidad Nacional del Centro de la Provincia de Buenos Aires-CICPBA-CONICET

mnatal@mdp.edu.ar

RESUMEN

Este trabajo surge del proyecto ImpaCT.AR N°38: “Intervenciones ambientales en el municipio de San Salvador de Jujuy: Modelado fluvial y monitoreo de una cuenca hídrica, biomonitoreo de la calidad del aire y restauración de áreas verdes”. La calidad del aire en San Salvador de Jujuy está afectada por las emisiones vehiculares y las industrias contaminantes de un municipio colindante. El material particulado (PM), especialmente $PM_{2.5}$ (partículas con diámetro aerodinámico $<2,5 \mu m$), es inhalable y puede afectar órganos vitales representando un riesgo para la salud y el ambiente. Se realizó la instalación de una red de sensores comerciales de bajo costo en trece sitios de la ciudad para registrar en tiempo real concentraciones de $PM_{2.5}$ y otras variables (PM_{10} , PM_1 , temperatura, presión atmosférica, humedad relativa). El objetivo de este trabajo es realizar un análisis espacial y temporal del $PM_{2.5}$ en el aire de la ciudad, identificando zonas potencialmente contaminadas y la estructura temporal. Los datos analizados corresponden a registros tomados cada 10 minutos de 13 sensores entre diciembre de 2023 y abril de 2024. Se aplicaron modelos geoestadísticos y de series de tiempo. Para el modelo geoestadístico, se realizó un análisis exploratorio, estructural (variograma empírico, ajuste a un modelo teórico y validación) y mapas de predicción mediante Kriging ordinario. Se aplicó la metodología de Box-Jenkins para estimar los modelos ARIMA que se ajustan a la serie temporal del $PM_{2.5}$ para cada sensor, y para los valores promedio de los trece sensores. Se detectaron dos zonas potencialmente contaminadas (Centro y barrio Alto Comedero) con valores altos de $PM_{2.5}$. Se obtuvo un modelo AR(1) correspondiente al promedio del $PM_{2.5}$ de los sensores. Debido a la naturaleza del problema, los resultados obtenidos de las series de tiempo serán fundamentales para la futura construcción de un modelo geoestadístico espacio-temporal.

Palabras clave: **predicción espacial - tendencia - estacionalidad - material particulado**



XXVIII Reunión Científica del Grupo Argentino de Bioestadística



Ciudad de Buenos Aires, octubre 2024



RIESGO RELATIVO DE INFECCIONES RESPIRATORIAS Y VIRUS SINCICIAL EN EL AMBA: UNA MIRADA DESDE LA ESTADÍSTICA ESPACIAL PARA ANALIZAR SU DINÁMICA PRE Y POSTPANDEMIA

DANIELA PARADA¹, SILVINA DEL DUCA^{1,2}

¹*Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

²*Departamento de Matemática, Facultad de Ciencias Económicas, Universidad de Buenos Aires*

daniela.parada@ic.fcen.uba.ar

RESUMEN

Las Infecciones Respiratorias Agudas Bajas (IRAB), que incluyen la neumonía y la bronquiolitis, son la principal causa de ingresos hospitalarios y de ingresos a la Unidad de Cuidados Intensivos Pediátricos (UCIP), a nivel mundial. Además, estas infecciones son las que más contribuyen al número de muertes de niños menores de 5 años. Varios patógenos respiratorios pueden causar las IRAB, siendo el Virus Sincicial Respiratorio (VSR) el agente etiológico predominante en la población de niños menores de cinco años. La carga de enfermedad por VSR impacta de manera sustancial en los servicios de salud. En este trabajo se presentan los resultados de un análisis exploratorio espacial de los datos de atención pediátrica de IRAB en cuatro hospitales públicos del Área Metropolitana de Buenos Aires (AMBA) entre los años 2018 y 2023. Tales datos fueron provistos por el equipo de investigadores del Centro INFANT de Medicina Traslacional (CIMeT) en el marco de un estudio sobre la dinámica de la infección por VSR en el contexto de COVID-19 en el AMBA, y de un estudio acerca de la correlación espacial de las IRAB severas y moderadas y el Índice de Vulnerabilidad Sanitaria (IVS). En particular, de nuestro trabajo se dependen la estimación de la intensidad de casos, tanto de IRAB como de VSR, así como la estimación espacial del riesgo relativo de la enfermedad por VSR severa para el período pre y postpandemia. Para todos los análisis se utilizó *R* (versión 4.2.3). Los resultados del trabajo aportan al actual estado del arte y pueden ser útiles para detectar oportunidades de intervención, tanto para los formuladores de políticas como para los profesionales de salud.

Palabras clave: ***datos georreferenciados, patógenos respiratorios, internación pediátrica, COVID-19.***

APLICACIÓN DE GAMM EN EL ESTUDIO DE LA RESPUESTA INMUNE HUMORAL POST VACUNACIÓN CONTRA COVID-19 EN ADULTOS MAYORES INSTITUCIONALIZADOS

TAMARA RICARDO¹, ANDREA SILVA¹, CHRISTIAN BALLEJO¹, M. FERNANDA AGUIRRE¹, M.
JIMENA MARRO¹, PAMELA E. RODRIGUEZ², PABLO RALL³, ESTEBAN A. MIGLIETTA², LUCÍA A.
LÓPEZ-MIRANDA¹, VERÓNICA PONCET¹, CARLA A. PASQUALE², YANINA MIRAGALLA³,
ANDREA GAMARNIK², ANDRÉS H. ROSSI²

¹Instituto Nacional de Epidemiología "Dr. J. H. Jara" (INE), ANLIS Malbrán, Mar del Plata,
Buenos Aires

²Fundación Instituto Leloir, CONICET, Buenos Aires

³Instituto Nacional de Servicios Sociales para Jubilados y Pensionados (INSSPP-PAMI), Buenos
Aires

tamararicardo@gmail.com

RESUMEN

Durante la pandemia de COVID-19 se observó alta incidencia de brotes de la enfermedad en establecimientos de estancia prolongada (EEP), favorecidos por la fragilidad de sus residentes y la permanencia en espacios semicerrados. Nuestro objetivo fue evaluar factores que influyeran sobre la respuesta inmune humoral en residentes de EEP de la provincia de Buenos Aires a 180 días post vacunación. Se realizó un estudio de cohorte multicéntrico con 773 voluntarios que recibieron dos dosis de vacunas Sputnik V, Sinopharm o AZD1222, analizando plasma a los 0, 21, 42, 120 y 180 días post primera dosis. Se evaluó el comportamiento de la concentración de anticuerpos IgG anti-Spike en el tiempo por grupo de exposición (naïve, no-naïve) y vacuna mediante modelos aditivos generalizados mixtos (GAMM), controlando dependencia temporal a partir de un término corCAR1 y efecto aleatorio del paciente, y la falta de independencia espacial por un intercepto aleatorio del EEP. Se analizaron como variables independientes la ocurrencia de brotes en el EEP y comorbilidades del/de la participante. Los valores de anticuerpos por debajo del límite de detección se imputaron usando un método de regresión lineal para datos censurados a la izquierda. El grupo naïve que recibió Sinopharm presentó niveles de anticuerpos significativamente menores que el resto ($p < 0,05$), mientras que en los no-naïve que recibieron Sinopharm se observaron niveles similares a los de los grupos naïve que recibieron Sputnik V y AZD1222. Quienes estuvieron expuestos a brotes en el EEP tuvieron niveles de anticuerpos significativamente mayores, independientemente del grupo de exposición y la vacuna ($P < 0,001$). La exposición previa, el tipo de vacuna y la pertenencia a un EEP con antecedente de brotes son factores a considerar frente a futuros eventos epidémicos de enfermedades infecciosas con dinámicas de transmisión similares al COVID-19.

Palabras clave: **COVID-19, adultos mayores, modelos aditivos generalizados mixtos (GAMM), dependencia temporal, datos longitudinales**

Ciudad de Buenos Aires, octubre 2024

ANÁLISIS DE MODELOS JERÁRQUICOS EN DATOS DE ESTRUCTURA ANIDADA UTILIZANDO DISTRIBUCIONES POISSON GENERALIZADAS

TKACH, KEVIN^{1,2} & FILLOY, JULIETA^{1,2}

¹*Laboratorio de Ecología de Comunidades y Macroecología, Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

²*Instituto de Ecología, Genética y Evolución (IEGEB), Universidad de Buenos Aires-CONICET*

tkachkevin@gmail.com

RESUMEN

Los conjuntos de datos organizados en estructuras jerárquicas pueden analizarse mediante modelos que permiten evaluar la variabilidad tanto dentro como entre los grupos. Examinamos la validez de los supuestos de cuatro modelos jerárquicos diferentes que asumen distribuciones de tipo Poisson según el diagnóstico de residuos obtenidos mediante simulación con el paquete DHARMA en R, para estudiar la relación entre las aves rapaces y las plantaciones forestales en paisajes donde preponderan coberturas de ambientes abiertos ubicados en el este de la Provincia de Entre Ríos (Región Pampeana). Inicialmente, se ajustó un modelo lineal generalizado con una distribución Poisson, utilizando la riqueza taxonómica como variable respuesta y la cobertura forestal (y su cuadrado) como predictores. El coeficiente de dispersión de Pearson obtenido fue mucho menor al supuesto (0.55 vs. 1), lo cual sugiere subdispersión de los residuos. Así, se probaron tres modelos alternativos para modelar la dispersión utilizando el paquete glmmTMB: con distribución Tweedie, Conway-Maxwell Poisson, y Poisson generalizado. Utilizando la misma estructura de factores fijos permitiendo una comparación directa, seleccionamos el modelo más adecuado según el criterio de Akaike o AIC y poniendo a prueba los supuestos de cada modelo. El modelo Poisson Generalizado resultó el más adecuado para capturar la estructura de la variabilidad en los datos jerárquicos analizados, considerando los valores de AIC y sus grados de libertad, bajo este modelo, la riqueza de rapaces aumentó hasta un máximo de 7 especies a 30% cobertura de plantaciones forestales, para luego descender a 1 especie a medida que la cobertura fue aumentando hasta alcanzar un 60% en el paisaje. Este estudio subraya la importancia de considerar modelos alternativos en el análisis de datos jerárquicos, especialmente cuando los supuestos de los modelos tradicionales no se cumplen.

Palabras clave: ***Ecología de comunidades, Homogeneización del paisaje, Cambio del uso del suelo, Aves rapaces***



Datos Longitudinales y Series Temporales

‘¿CÓMO HARÍAS PARA SABER CUÁNTO PIS HACE UNA PERSONA EN UN DÍA?’ DESARROLLO DE HABILIDADES DE PENSAMIENTO CIENTÍFICO EN ADOLESCENTES ARGENTINOS

MARTINA BOSCOLO¹, MELINA FURMAN^{2,3}, LORENA RELA⁴, PEDRO BEKINSCHTEIN⁵, RODRIGO LAJE^{2,6},
ANDREA P. GOLDIN¹

¹ *Laboratorio de Neurociencia. Centro de Inteligencia Artificial y Neurociencia (CIAN). Escuela de Negocios. Universidad Torcuato Di Tella. CONICET*

² *CONICET, Buenos Aires.*

³ *Escuela de Educación, Universidad de San Andrés.*

⁴ *Instituto de Fisiología y Biofísica Bernardo Houssay, Facultad de Medicina, Universidad de Buenos Aires, CONICET.*

⁵ *Laboratorio de Neurociencia y Sociedad, Instituto de Neurociencia Cognitiva y Traslacional, CONICET, Fundación INECO, Universidad Favaloro.*

⁶ *Laboratorio de Dinámica Sensoriomotora, Departamento de Ciencia y Tecnología, Universidad Nacional de Quilmes, Bernal
martiboscolo@gmail.com*

RESUMEN

El pensamiento científico es un conjunto de procesos cognitivos intencionales de búsqueda y aplicación de conocimiento que implica plantear preguntas investigables, diseñar formas de responderlas, y usar evidencia para revisar y modificar el propio conocimiento. Es una de las habilidades esenciales para el siglo XXI. Los campamentos científicos son comunes en la educación científica informal de países centrales, pero no así en América Latina. Durante 8 años recogimos datos de adolescentes argentinos que participaron en uno de los ocho campamentos científicos independientes de una semana de duración de la Asociación Civil *Expedición Ciencia*. El objetivo del proyecto fue determinar si los campamentos mejoraban ciertas habilidades de pensamiento científico, y si el cambio dependía del nivel socioeconómico (NSE). Se evaluó el rendimiento de 290 adolescentes de entre 13 y 18 años a través de las respuestas a un instrumento administrado antes y después de la intervención. Se realizó un modelo mixto de enlace acumulativo (CLMM) que incluyó Puntaje como la variable dependiente; Prueba (pretest/posttest), NSE (alto/bajo) y Pregunta (niveles correspondientes a las 4 preguntas del instrumento) como variables explicativas; las interacciones entre ellas; y Participante como variable de efectos aleatorios. Además, se utilizaron métodos de *Bootstrapping* para determinar si la mejora de la intervención difería significativamente de 0. Se encontró que el desempeño después de la intervención fue significativamente mayor que antes y, notablemente, el efecto de la intervención no dependía del nivel NSE. Además, los adolescentes de mayor nivel NSE se desempeñaron mejor que sus pares de menor NSE tanto antes como después de la intervención. Se presentan métodos estadísticos útiles y relativamente poco utilizados en la Psicología Cognitiva y Experimental que permiten evidenciar que ciertas habilidades de pensamiento científico pueden entrenarse y enseñarse con solo una semana de entrenamiento.

Palabras clave: ***pensamiento crítico, transferencia, entrenamiento, adolescencia, nivel socioeconómico.***

Ciudad de Buenos Aires, octubre 2024

ANÁLISIS DE LOS EFECTOS DEL HIPOTIROIDISMO EN DIFERENTES ETAPAS DE LA VIDA SOBRE EL CRECIMIENTO Y PERFIL HORMONAL

MARÍA JOSÉ CASTRO¹, KIARA PELOSO², EUGENIA PETRONAVE², SOFIA EPÍSCOPO², ANDREA FELLET², MYRIAM NUÑEZ¹

¹ *Cátedra de Matemática, Departamento de Fisicomatemática, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires*

² *Cátedra de Fisiología, Departamento de Ciencias Biológicas, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires*

mjcastro@ffyb.uba.ar

RESUMEN

Las hormonas tiroideas tienen una acción muy importante en la embriogénesis y maduración del feto. Estas hormonas son esenciales para el desarrollo, proliferación y maduración del sistema nervioso central y de otros tejidos. Teniendo en cuenta estos antecedentes, el objetivo del estudio fue explorar si el hipotiroidismo inducido durante diferentes etapas de la vida (período intrauterino, lactancia y perinatal) afecta el crecimiento, así como también al sistema cardiovascular y estado redox (procesos de oxidación y reducción) en la adultez. Las pruebas fueron realizadas en ratas Sprague-Dawley las que se dividieron en cuatro grupos: G (crías macho de madre que tenía acceso libre al agua con 0.02% de metimazol (MMZ) durante la gestación), GL (crías macho de madre que tenía acceso libre al agua con 0.02% de MMZ durante la gestación y lactancia), C (crías macho de madre que tenía acceso libre al agua corriente) y D (crías macho tratados con MMZ de madre que tenían acceso libre al agua corriente). Las principales variables estudiadas fueron: peso, longitud de la cola, longitud de la tibia, tiroxina (T4), triyodotironina (T3), y hormona estimulante de la tiroides (TSH). Las tres primeras variables fueron medidas a lo largo del tiempo. Como primer paso se realizó un análisis descriptivo de las variables en estudio. Con el fin de evaluar la existencia de diferencias significativas entre los grupos anteriormente mencionados se aplicó la prueba de Kruskal Wallis y la correspondiente prueba a posteriori y ANOVA de dos factores con medidas repetidas en un factor según correspondiera. Este estudio sugiere que el hipotiroidismo inducido en las etapas tempranas de la vida condicionaría la actividad redox, explicando así el deterioro y los cambios estructurales observados. Esto subraya la necesidad de comprender y analizar los impactos derivados de las alteraciones tiroideas siendo que podrían generar repercusiones duraderas a lo largo de la vida mediante modificaciones permanentes en la morfología.

Palabras clave: ***anova, hipotiroidismo, actividad redox.***

SALUD DE LAS COLONIAS DE ABEJAS EN SISTEMAS APÍCOLAS SANTAFESINOS DE PEQUEÑA ESCALA

FURLAN EXEQUIEL¹; MAGGI MATÍAS^{2,3}, BELDOMENICO PABLO¹, JUNGES CELINA¹

¹Laboratorio de Ecología de Enfermedades, Instituto de Ciencias Veterinarias del Litoral (ICiVet-Litoral), Universidad Nacional del Litoral - CONICET, Esperanza, Santa Fe

²Centro de Investigación en Abejas Sociales (CIAS), Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata, Mar del Plata, Buenos Aires

³Instituto de Investigaciones en Producción Sanidad y Ambiente (IIPROSAM- CONICET-UNMdP), Mar del Plata

exequielfurlan@icivet.unl.edu.ar

RESUMEN

La apicultura en Argentina, y en el mundo, enfrenta desafíos como la pérdida de hábitats, el uso de pesticidas y la presión de patógenos y parásitos, que afectan la salud y supervivencia de las colonias de abejas (*Apis mellifera*). Este estudio analiza cómo los ácaros de *Varroa* y los esporos de *Vairimorpha* spp. influyen en el estado de salud de las colonias manejadas por pequeños productores (\leq de 50 colmenas) del centro de Santa Fe. Se llevó a cabo un estudio longitudinal, con 45 colmenas seleccionadas al azar de nueve apiarios diferentes, muestreadas en cuatro momentos: dos en la temporada productiva y dos en la temporada no productiva. En cada visita, se cuantificaron el número de ácaros y esporos, y las proporciones de polen, cría abierta, cría cerrada y miel como indicadores de salud de las colonias. Se utilizaron modelos lineales mixtos generalizados (GLMM) y ANOVA para relacionar estas variables con la temporalidad y el grado de infestación de los parásitos. Los GLMM revelaron que: I) La proporción de polen varió significativamente con la interacción entre los niveles de *Varroa* y las visitas ($X^2 = 13,04$; $p = 0,042$); II) La proporción de cría abierta mostró variaciones entre visitas ($X^2 = 125,94$; $p < 0,001$) y con los niveles de *Varroa* ($X^2 = 6,73$; $p = 0,035$); III) tanto la proporción de cría cerrada ($X^2 = 137,24$; $p < 0,001$) como la cantidad de miel almacenada ($X^2 = 28,28$; $p < 0,001$) variaron significativamente entre visitas. Los hallazgos sugieren que la estacionalidad es el factor principal en la variación de los indicadores de salud, aunque la dinámica de los ácaros también influye en la producción de polen y la cría abierta, subrayando la importancia de considerar ambos factores para una gestión efectiva de la salud de las colonias.

Palabras clave: *Apis mellifera*, salud, parásitos



MODELOS LINEALES GENERALIZADOS MIXTOS PARA COMPARAR ANTIPARASITARIOS DESDE EL CONTEO DE NEMATODOS EN OVEJAS CRIOLLAS

ROVEGNO, MS¹, PEÑA, S¹, LOPEZ, G¹ ABBIATI, NN¹

¹ Instituto de Investigación sobre la Producción Agropecuaria, Ambiente y Salud (IIPAAS), Facultad de Ciencias Agrarias, Universidad Nacional de Lomas de Zamora

mrovegno@agrarias.unlz.edu.ar

RESUMEN

Dentro de los endoparásitos ovinos presentes en la Argentina, los nematodos son los de mayor importancia, debido a los problemas de salud y pérdidas económicas que ocasionan. El objetivo fue comparar dos antiparasitarios (tradicional, Homeopático) y un Control; evaluándolos a través del conteo de huevos por gramos de materia fecal (HPG) a lo largo del tiempo. El diseño experimental fue completamente aleatorizado con 58 ovejas criollas, 20 para el tratamiento Homeopático, 19 para Albendazol (tradicional) y Control. Se efectuaron 4 mediciones, al momento 1 donde se aplicaron los antiparasitarios y luego a los 15, 29 y 50 días posteriores. Se emplearon los softwares SAS y Navure, con $\alpha=0.05$ y comparaciones múltiples de Tukey. Dado que la variable HPG es un recuento con mayor varianza que media, se la analizó mediante modelos lineales generalizados mixtos (MLGM), contemplando como media el tiempo, los antiparasitarios y su interacción. Se probaron diferentes distribuciones, siendo la binomial negativa con función de ligadura logarítmica y correlación a lo largo del tiempo con simetría compuesta, con el efecto del animal como aleatorio, la que mejor ajustó, para ambos software. Ambos programas presentaron resultados similares, detectando interacción entre antiparasitario y tiempo (valor- $p < 0.0001$). SAS permitió comparar las medias de los antiparasitarios dentro de cada tiempo y Navure lo hizo sobre las medias de las interacciones. Al momento 1 no se detectaron diferencias, como era esperable. En el momento 15 se diferenció solamente el Albendazol, mientras que en los otros momentos todos fueron diferentes, siendo el control el que presentó mayor HPG promedio y Albendazol el menor. Es posible concluir que no se diferenciaron los resultados obtenidos a través de los softwares utilizados en este ensayo, aunque SAS permite comparar promedios dentro de un factor cuando hay interacción de dos factores.

Palabras clave: **modelos lineales generalizados mixtos, ovinos, antiparasitarios, SAS, Navure.**



MÉTRICAS FENOLÓGICAS DEL BOSQUE DE LENGAS EN EL NO DE CHUBUT UTILIZANDO DIFERENTES ESCALAS TEMPORALES DE DATOS DE NDVI-MODIS

ROCÍO M. DE LELLIS¹, PIEDAD M. CRISTIANO¹², PAULA I. CAMPANELLO³

¹Laboratorio de Ecología Funcional, Departamento de Ecología Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.

²Instituto de Ecología, Genética y Evolución (IEGEB-CONICET), Universidad de Buenos Aires

³Instituto de Biotecnología Esquel, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco. Esquel, Chubut, CONICET.

rociodelellis@gmail.com

RESUMEN

La fenología de la vegetación, como la brotación y caída de hojas, influye en la productividad primaria y en los ciclos de agua, nutrientes y carbono a nivel ecosistémico. En climas templados, la primavera aumenta la temperatura y provoca brotación y crecimiento, mientras que en otoño, la disminución de temperatura y luz induce la caída de hojas. En zonas montañosas, la temperatura y el deshielo son claves para la fenología de las plantas, con impactos más fuertes a mayor altitud. Debido al cambio climático, la fenología ha cambiado significativamente desde mediados del siglo pasado, especialmente en los últimos 40 años. Sin embargo, las observaciones climáticas y fenológicas son prácticamente inexistentes en los bosques andinos del sur. El objetivo de este trabajo fue analizar las métricas fenológicas del bosque de lengas del noroeste de la provincia de Chubut a diferentes altitudes (1200, 1300 y 1400 metros) utilizando productos de NDVI (Normalized Difference Vegetation Index) del sensor MODIS (Moderate Resolution Imaging Spectroradiometer) con dos resoluciones temporales (8 y 16 días), mediante el paquete phenofit disponible para R. Para ello se ajustó un modelo a cada serie temporal del NDVI para capturar las señales estacionales de la vegetación, que posteriormente se utilizaron para identificar los períodos de crecimiento (PC). Luego se aplicó otro ajuste para capturar los cambios rápidos de la vegetación, utilizado para extraer las métricas fenológicas correspondientes a cada PC mediante el método de inflexión: reverdecimiento, madurez, senescencia y dormición. Se logró una mayor detección de los PC en las series temporales con una resolución de 16 días y en puntos ubicados a mayor altura, que en las series con una resolución temporal de 8 días y en puntos ubicados a menor altura. A su vez, la cantidad de PC disminuyó con la cantidad de iteraciones de cada ajuste.

Palabras clave: **fenología, gradiente altitudinal, Nothofagus pumilio, paquete phenofit, R**



Educación estadística



INTEGRACIÓN DEL SOFTWARE R, RSTUDIO Y POSIT CLOUD EN LA EDUCACIÓN ESTADÍSTICA: EXPERIENCIA CON ESTUDIANTES DE INGENIERÍA AGRONÓMICA

BUSTOS, EDUARDO SEBASTIAN¹, ROMERO ARNALDO ANTONIO¹, RODRÍGUEZ NORMA LEONOR¹

¹*Facultad de Ciencias Agrarias, Universidad Nacional de Catamarca*

sbustos@agrarias.unca.edu.ar

RESUMEN

La enseñanza estadística busca mejorar el proceso de aprendizaje. La Cátedra de Biometría de la Facultad de Ciencias Agrarias de Catamarca promueve un cambio de enfoque hacia la comprensión conceptual, el pensamiento estadístico y el aprendizaje activo, empleando herramientas didácticas y tecnológicas. Este trabajo presenta una experiencia educativa con estudiantes de Ingeniería Agronómica, donde se utilizó el software R, RStudio y Posit Cloud, según la percepción de los estudiantes que cursaron Estadística y Biometría en el segundo año de la carrera. Se realizó un estudio descriptivo transversal con 18 estudiantes. Para ello, se diseñó una encuesta con preguntas cerradas, semiabiertas y abiertas, administrada a través de Google Forms. Los datos se analizaron estadísticamente con el software R. La muestra estaba integrada por ocho hombres y diez mujeres, con edad promedio de 22,3 años. El 77,8% no tenía conocimientos previos de R, mientras que el 22,2% tenía un nivel básico. El 44,4% consideró al software extremadamente útil, el 50,0% útil y el 5,6% moderadamente útil. El 61,1% encontró el nivel de dificultad fácil a normal, mientras que el 38,9% lo percibió como difícil. Según los estudiantes, los principales desafíos fueron identificar y corregir errores en el código, la curva de aprendizaje inicial y los retos para crear y usar scripts. Las ventajas marcadas fueron la creación de gráficos, capacidad para análisis avanzados y los paquetes adicionales. Respecto al uso de Posit Cloud, destacaron la rapidez, facilidad de uso y accesibilidad desde múltiples dispositivos, con la limitación de tiempo de uso disponible como principal desventaja. La mayoría de los estudiantes (94,4%) recomienda continuar enseñando la asignatura utilizando estas herramientas. Los resultados indican una percepción positiva hacia el uso de estas tecnologías, sin embargo, la curva de aprendizaje y las limitaciones de tiempo en Posit Cloud requieren atención para optimizar la experiencia educativa.

Palabras clave: *enseñanza estadística, software R, RStudio, Posit Cloud, aprendizaje activo.*



INNOVACIÓN EN LA ENSEÑANZA DE ESTADÍSTICA EN CIENCIAS AGRARIAS: INTEGRACIÓN DE R Y RSTUDIO EN ENTORNOS EDUCATIVOS

PATRICIA MELINA ACETTA¹, CARLOS AGUSTIN ALESSO², ALEJANDRO ALEMÁN¹

¹*Facultad de Ciencias Agrarias, Universidad Nacional del Litoral*

²*CiAgro Litoral, UNL-CONICET, Facultad de Ciencias Agrarias, Universidad Nacional del Litoral*

pacetta@fca.unl.edu.ar

RESUMEN

Este trabajo explora la integración de R y RStudio en la enseñanza de la estadística en cursos de la carrera de grado Ingeniería Agronómica y cursos de posgrado, destacando su impacto en el aprendizaje de conceptos estadísticos. Se presenta la importancia de utilizar herramientas modernas como R y RStudio para la enseñanza de la estadística, mejorando la comprensión teórica y práctica, a través de la utilización de RStudio Desktop y RCloud en cursos de Estadística I y II, Introducción al Análisis de Datos con R y Diseño Experimental. Se analiza cómo RStudio Desktop se emplea para la enseñanza de estadísticas descriptivas, probabilidad y análisis de datos, mientras que RCloud facilita la colaboración en línea y el análisis en la nube. Se describen los trabajos prácticos y su conexión con la teoría estadística en estos cursos. Los resultados muestran una mejora en la comprensión de los conceptos estadísticos por parte de los estudiantes que utilizaron estas herramientas, evidenciando una mayor aplicación práctica y visualización de datos. La discusión aborda los beneficios, como el enriquecimiento de la experiencia de aprendizaje y la capacidad de realizar análisis complejos, así como los desafíos, incluyendo la curva de aprendizaje y el soporte técnico necesario. Finalmente, se presentan propuestas para la mejora continua en la integración de estas herramientas en el currículo académico, considerando su evolución y el potencial impacto en la enseñanza de la estadística. Este análisis destaca cómo el uso de R y RStudio, tanto en versiones de escritorio como en la nube, ofrece una metodología efectiva para la enseñanza y aprendizaje de la estadística, promoviendo una comprensión más profunda y práctica de los conceptos. Se proponen estrategias para continuar mejorando la integración de estas herramientas en el currículum académico, adaptándose a su evolución y potenciando su impacto en la enseñanza de la estadística.

Palabras clave: ***aprendizaje, metodología educativa, enseñanza de la estadística con R, análisis de datos para ingeniería***



ANÁLISIS DE LOS SABERES PREVIOS DE LOS ESTUDIANTES DE BIOESTADÍSTICA

MYRIAM NUÑEZ¹, MARÍA EUGENÍA BONADIE¹, PAULA ZAMBIANCHI¹, MATÍAS CAMALET-LE NOBLE¹,
JUDITH MONTENEGRO BRUSOTTI¹

¹ *Cátedra de Matemática, Departamento de Fisicomatemática, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires*

myriam@ffyb.uba.ar

RESUMEN

El objetivo de este trabajo es realizar un análisis cuantitativo que permita describir los conocimientos previos de los estudiantes del primer cuatrimestre de 2024 de la asignatura Bioestadística de las carreras de Farmacia y Bioquímica, respecto de la resolución de la Prueba Diagnóstica (PD) y compararlo con resultados previos (pre-pandemia, primer cuatrimestre de 2019 y post-pandemia, primer cuatrimestre de 2022). La PD se estructura como un cuestionario de diez ejercicios, clasificados según las categorías: Operaciones Algebraicas (OA), Cálculo Porcentual (CP), Derivadas e Integrales (DI), Cálculo de Áreas (CA) y Teoría de Conjuntos (TC), contenidos necesarios para la cursada de la asignatura. Se incluye la nota de Matemática ya que es correlativa de Bioestadística. De los resultados obtenidos se puede resaltar que el ejercicio de inecuaciones, categoría OA, presenta el menor porcentaje de respuestas correctas, seguido del de integrales definidas de la categoría DI. La condición de haber aprobado Matemática está asociada a la respuesta correcta de los ejercicios, tanto en 2019 ($p = 0,0475$), como en 2022 ($p = 0,0159$), no así en 2024 ($p = 0,1228$). En el caso del ejercicio de integral definida sólo se verificó asociación entre el hecho de haber aprobado Matemática y la correcta resolución de este en 2019 ($p = 0,0208$). Para el ejercicio de inecuación, no se halló asociación para ninguno de los años analizados. En conclusión, si bien en la mayoría de las categorías no volvimos a los resultados del 2019, logramos superar los resultados obtenidos en 2022. Esto es tenido en cuenta para diseñar estrategias que favorezcan la apropiación de dichos contenidos como hacer hincapié en los temas correspondientes a las categorías anteriormente mencionadas que se retoman en las clases de consulta que imparte la cátedra porque son necesarios para la resolución de los ejercicios que se plantean en la práctica.

Palabras clave: ***bioestadística, saberes previos, prueba diagnóstica, categorías.***



EVALUACIÓN DE LOS CONOCIMIENTOS PREVIOS DE LOS ESTUDIANTES DE ÓPTICA GEOMÉTRICA LUEGO DEL CAMBIO DE PLAN DE ESTUDIOS

JUDITH MONTENEGRO BRUSOTTI^{1, 2}, PAULA ZAMBIANCHI¹, MATÍAS CAMALET-LE NOBLE¹, MYRIAM NUÑEZ¹

¹Cátedra de Matemática, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires

² Cátedra de Óptica, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires

jmontenegro@ffyb.uba.ar

RESUMEN

El objetivo de este trabajo es analizar si los conocimientos previos de los estudiantes de la asignatura Óptica Geométrica (OG) de la Tecnicatura Universitaria en Óptica y Contactología de la Facultad de Farmacia y Bioquímica de la UBA, han sido nivelados a partir de los cambios en el Plan de Estudios. Una de las modificaciones realizadas es la incorporación de materias del Ciclo Básico Común (CBC), en particular Matemática 51. Para llevar a cabo este objetivo se utilizó una prueba diagnóstica (PD), al inicio de la cursada, clasificando los ejercicios en tres grupos por bloque temático (G_1 : Matemática, G_2 : Física, G_3 : Geometría). Se consideraron los resultados correspondientes de los estudiantes de OG que realizaron la PD en 2024 ($n = 66$) y los que pertenecían al plan anterior que cursaron en 2023 ($n = 59$), según las tres categorías. Además, se efectuó una partición de los resultados de la PD en tres grupos, considerando el puntaje de los ejercicios bien resueltos (P_1 : menor o igual a 2, P_2 : igual a 3, P_3 : entre 4 y 5). Se realizó un análisis descriptivo de las variables intervinientes y, con el fin de evidenciar si existen diferencias a partir de la modificación efectuada, se compararon los grupos aplicando la prueba de diferencia de proporciones y la Prueba de Ji-Cuadrado para estudiar el grado de asociación entre las variables. Se evidenció una mayor dificultad en el bloque temático G_2 . Los resultados obtenidos en 2023 fueron mejores que los de 2024. Esto nos obliga a repensar en una asignatura del CBC que sea propia para los estudiantes de esta tecnicatura. En tanto a este cuatrimestre, se generarán estrategias de enseñanza que se adecuen a los conocimientos previos sin disminuir la exigencia académica y lograr mayor permanencia de los estudiantes en el ámbito educativo.

Palabras clave: **conocimientos previos, plan de estudios, estrategias de enseñanza, óptica.**



VISUALIZACIÓN DE LA SITUACIÓN ARGENTINA A PARTIR DE INDICADORES DE ACCESO PÚBLICO

RUMINOT SERGIO¹, RUBIO NATALIA¹, SABINO GUILLERMO¹, CARDELLI BRUNO¹, GIMENEZ GUSTAVO N.¹

¹ *Departamento de Estadística, Facultad de Economía, Universidad Nacional del Comahue*

natalia.rubio@faeauncoma.com.ar

RESUMEN

El acceso a bases de datos de indicadores socio-económicos de los países se agiliza día a día. El uso de librerías, API (application programming interface) y la descarga de archivos planos de gran volumen de información permiten a los analistas de datos conocer el estado de situación pasado y actual de los países. Sin embargo, el preprocesamiento de esas bases, constituye una de las principales barreras a derribar si se desea visualizar y analizar dicha información. Los objetivos de este trabajo han sido establecer los pasos necesarios para obtener las bases de datos que ayuden por un lado a describir la situación pasada y actual de la Argentina en cuanto a la evolución del Índice de Desarrollo Humano (HDI) y por otro evaluar la inflación de los últimos treinta años. Los datos del HDI se obtuvieron a partir de un archivo CSV del Programa de Desarrollo de las Naciones Unidas, para describir la inflación se trabajó con datos del INDEC, y se accedió a algunos indicadores de pobreza del Banco Mundial mediante la librería *wbstats*. A partir de un trabajo reproducible en R se reordenaron las bases de formato ancho a largo, se eliminaron, unificaron y renombraron variables mediante las funciones de las librerías *dplyr* y *tidyr*. Se realizaron visualizaciones comparando la Argentina en el contexto Sudamericano y mundial con respecto al HDI, y se comenzó una visualización que proporciona posibles tendencias futuras de la inflación en Argentina. Entre los principales resultados se visualiza que Argentina tiene un HDI superior al resto de Sudamérica, excepto entre 2016-2022 cuando Chile lo superó. Los gráficos de los pronósticos del modelo ARIMA destacaron que este último año es especialmente preocupante por la alta variabilidad de los últimos tiempos. Se espera arribar a un modelo predictivo final que mejore las estimaciones gráficas preeliminares.

Palabras clave: *Índice de Desarrollo Humano, Inflación en Argentina*



DISTRIBUCIÓN GEOGRÁFICA Y ANÁLISIS SOCIOECONÓMICO DE LOS INGRESANTES A LA FACULTAD DE ECONOMÍA Y ADMINISTRACIÓN (FAEA) DE LA UNIVERSIDAD NACIONAL DEL COMAHUE, 2024

NATALIA RUBIO¹, SERGIO RUMINOT² GUSTAVO N GIMENEZ³

¹ *Departamento de Estadística, Facultad de Economía, Universidad Nacional del Comahue*

natalia.rubio@faeauncoma.com.ar

RESUMEN

Desde el área de ingreso y permanencia de la FAEA se realizó un relevamiento llevado a cabo por los tutores pares a un total de 278 ingresantes, teniendo en cuenta sus principales características socioeconómicas. En el marco del análisis, se realizó un estudio descriptivo utilizando diferentes representaciones gráficas como mapas de calor sobre el mapa de Argentina y gráficos interactivos que muestran las diferentes variables estudiadas. Se registraron estudiantes provenientes de 16 provincias del país, con su mayor porcentaje en Neuquén (58,3%), seguido por la provincia de Río Negro (25,5%) y Buenos Aires (4,68%). La carrera con mayor cantidad de inscriptos es Contador Público Nacional (CPN), con un 43.5%, y las de menor cantidad son Profesorado de Ciencias Económicas y Ciclo General de Ciencias Económicas, con un 0.7%. Los ingresantes asisten a la facultad diariamente desde un total de 9 localidades de Neuquén y Río Negro. Del 43,5% de los estudiantes de CPN, un 53,7% residen en la ciudad de Neuquén Capital, un 10,7% en la ciudad de Centenario, un 8,26% en la Ciudad de General Roca a 46 km de la FAEA y un 0,8% de Ingeniero Huergo a 77 km respectivamente. El análisis permitió identificar a los estudiantes que residen en ciudades más alejadas, como Ingeniero Huergo y General Roca, y que se trasladan periódicamente a la FAEA. Esta información facilita una planificación más efectiva de la asignación de recursos y la adaptación de las políticas institucionales para comprender mejor las necesidades de los estudiantes. No solo permite repensar los programas de apoyo, así como las becas de residencia y económicas, sino que también resalta la importancia del carácter gratuito de la educación universitaria en Argentina, que facilita el acceso de estudiantes provenientes de diversas partes del país.

Palabras clave: **procedencia geográfica, perfil socioeconómico, inclusión educativa.**



APRENDIZAJE AUTODIRIGIDO: CLAVE PARA EL DESARROLLO DE HABILIDADES DE PENSAMIENTO ESTADÍSTICO

MARÍA CRISTINA SANCHEZ FIGUEROA¹, GUILLERMO SABINO², LILIANA TAUBER³

¹Facultad de Ciencias Económicas y Empresariales, Universidad Nacional de Educación a Distancia

²Facultad de Economía y Administración, Universidad Nacional del Comahue, Neuquén

³Facultad de Humanidades y Ciencias, Universidad Nacional del Litoral, Santa Fe

guillermo.sabino@faea.uncoma.edu.ar

RESUMEN

Debido al incremento exponencial del volumen de datos, visualizar y comunicar de manera efectiva la información subyacente emerge como una nueva asignatura en los programas académicos de las carreras de grado. En base a la diversidad de intereses por parte de los estudiantes, el aprendizaje autodirigido juega un papel esencial en el desarrollo de habilidades en análisis de datos y visualización. Este enfoque permite a los estudiantes identificar sus propias necesidades de aprendizaje, establecer objetivos claros y seleccionar los recursos más adecuados para alcanzar esos objetivos. La autodirección en el aprendizaje fomenta una actitud proactiva y un compromiso más profundo con el proceso educativo, resultando en una comprensión más sólida y aplicada de los conceptos y técnicas. Por otra parte, la internacionalización en la educación ha ido cobrando una importancia cada vez mayor y representa, en cuanto a calidad y equidad, un reto para el sistema de educación superior. Estos tres ejes – la visualización de datos (dataviz), el aprendizaje autodirigido, la internacionalización online – ha propiciado la colaboración conjunta de instituciones educativas de diversos países de habla hispana. En el marco del *Collaborative Online International Learning (COIL)*, se presenta una propuesta que engloba un nuevo paradigma de enseñanza y aprendizaje, la cual desarrolla la conciencia intercultural en entornos de aprendizaje compartidos entre estudiantes de la Universidad Nacional de Educación a Distancia (España) y de la Universidad Nacional del Comahue (Argentina). Las actividades propuestas por los docentes como facilitadores del aprendizaje están diseñadas para desarrollar habilidades de pensamiento estadístico: el estudiante accede a bases de datos de su propio interés de diferentes maneras (data wrapping, APIs, etc.), discute críticamente las potencialidades y falencias de gráficos hallados en la web, genera visualizaciones de interés paso a paso y elabora una producción final en grupo, acompañado de una comunicación acorde de sus resultados.

Palabras clave: ***Aprendizaje Autodirigido, Dataviz, Habilidades de Pensamiento Estadístico, Educación Estadística en el Nivel Superior***

CUALIDADES DE LOS INSTRUMENTOS PARA LA EVALUACIÓN SUMATIVA DE CONCEPTOS ESTADÍSTICOS APLICADOS EN PROBLEMAS DE CIENCIAS AGROPECUARIAS

FRANCO SUAREZ¹, FABRIZIO GARCIA SELEME², DAVID MORALES¹, MARIANO CÓRDOBA^{1,2},
PABLO PACCIORETTI^{1,2,3}, MARGOT TABLADA¹, CECILIA BRUNO^{1,2}

¹*Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba*

²*Grupo vinculado de Estadística y Biometría Unidad de Fitopatología y Modelización Agrícola (UFyMA-INTA-CONICET), Córdoba*

³*Departamento de Agronomía y Horticultura, Universidad de Nebraska*

suarezfranco@agro.unc.edu.ar

RESUMEN

En la evaluación educativa es fundamental contar con instrumentos de medición precisos y confiables para tomar decisiones que fortalezcan el logro de competencias. La psicometría permite aplicar modelos estadísticos a las respuestas obtenidas en una evaluación, para determinar la confiabilidad del instrumento utilizado. La Teoría de Respuesta al Ítem modela la probabilidad de respuesta correcta a una pregunta (ítem) en función de un conjunto de parámetros que describen tanto las características del ítem como las habilidades del evaluado. El objetivo de este trabajo fue analizar las cualidades de las preguntas utilizadas en evaluaciones sumativas de la asignatura Estadística y Biometría en las carreras de la Facultad de Ciencias Agropecuarias de la Universidad Nacional de Córdoba. Se utilizaron datos de las evaluaciones de suficiencia (ES) 1 y 2 de la cohorte 2024. Los temas incluidos en la ES1 fueron estadística descriptiva, probabilidad, distribución de estadísticos muestrales y comparación de dos poblaciones con un total de 29 preguntas completadas por 202 estudiantes. La ES2 abordó el diseño y análisis de experimentos a uno y dos criterios de clasificación y la regresión lineal simple, con 30 preguntas respondidas por 183 estudiantes. Se ajustó un modelo logístico de tres parámetros: discriminación (a); dificultad (b) y azar (c). Los ítems fueron clasificados en cuatro: teóricos, de cálculo, identificación de resultado e interpretación de resultados en el contexto del problema. Los ítems de cálculo presentaron, en promedio, la mayor capacidad para discriminar entre estudiantes con diferentes niveles de habilidad ($a=2.63$), seguidos por los ítems de interpretación ($a=2.00$). Los ítems de identificación de resultados resultaron, en promedio, los más fáciles ($b=-0.27$), mientras que los de interpretación los más difíciles ($b=0.22$). Los instrumentos distinguieron, sobre todo, niveles de habilidad en el uso de conceptos y resultados estadísticos para obtener información contextualizada.

Palabras clave: **competencias, teoría respuesta al ítem, modelo logístico, discriminación.**



Meta-Análisis

META-ANÁLISIS COMO HERRAMIENTA PARA LA EVALUACIÓN DE RESPUESTAS DE AVES AL MANEJO GANADERO: UN ESTUDIO DE CASO EN LOS PASTIZALES DEL RÍO DE LA PLATA

FACUNDO NIKLISON^{1,2}, DAVID BILENCA^{1,2}, MARIANO CODESIDO^{1,2}

¹*Instituto de Ecología, Genética y Evolución de Buenos Aires. Universidad de Buenos Aires - CONICET*

²*Grupo de Estudios de Biodiversidad en Agroecosistemas*

facuniklison@gmail.com

RESUMEN

Los meta-análisis surgieron en las ciencias sociales hacia la década de 1970, en tanto que su primera aplicación en ecología data de 1991. Si bien desde entonces el uso de revisiones sistemáticas que incluyen meta-análisis se expandió en nuestra disciplina, surgieron en paralelo autores que señalan tanto sus ventajas como algunas limitaciones. En este trabajo consideramos algunos de estos aspectos aplicados a un estudio de caso en el que analizamos mediante una revisión sistemática (con meta-análisis) las respuestas del ensamble de aves de pastizal a distintos manejos ganaderos en los Pastizales del Río de la Plata. Al respecto, encontramos que: 1) el meta-análisis reduce el riesgo de tomar decisiones erradas basadas en un caso de estudio que puede no seguir la tendencia global del sistema, 2) pudimos incluir como moderador a una variable cuya influencia es difícil de evaluar a escala de un único estudio (en este caso, la altura del pastizal), lo que nos permitió explicar parte de la heterogeneidad hallada en las magnitudes de efecto, 3) las conclusiones de nuestros resultados se vieron limitadas por la baja cantidad de estudios con datos accesibles, lo que resalta la importancia de informar la estadística básica en las publicaciones (tamaño de muestra, medidas de tendencia central y de dispersión) 4) Otra limitación de los meta-análisis (que también ocurre en las revisiones cualitativas) es la posibilidad de sesgo de publicación, aunque en nuestro caso lo evaluamos estadísticamente y no lo detectamos, y 5) un modelo de efectos aleatorios nos permitió controlar la falta de independencia entre datos provenientes de un mismo artículo. Nuestro trabajo resalta el aporte del meta-análisis como herramienta cuantitativa para estudiar y evaluar alternativas de manejo, siempre y cuando se tomen en consideración las limitaciones mencionadas y se les otorgue un tratamiento e interpretación adecuados.

Palabras clave: *metodologías cuantitativas, meta-análisis, revisiones ecológicas*

DESAFÍOS METODOLÓGICOS EN UNA REVISIÓN SISTEMÁTICA VIVA Y META-ANÁLISIS INTERACTIVOS QUE EVALÚA LA SEGURIDAD Y EFICACIA DE VACUNAS COVID 19 EN EMBARAZADAS

NOELIA CASTELLANA¹, DIEGO MARFETAN MOLINA², JAMILE BALLIVAN¹, MABEL BERRUETA¹,
AGUSTIN CIAPPONI¹

¹*Instituto de Efectividad Clínica y Sanitaria (IECS)*

²*Universidad Nacional de Rosario*

ncastellana@iecs.org.ar

RESUMEN

Con el objetivo de evaluar los beneficios y riesgos de las vacunas contra el COVID-19 durante el embarazo, se desarrolló una plataforma web interactiva que permite visualizar los resultados actualizados de una revisión sistemática viva y meta-análisis realizada sobre el tema. Financiado por la Fundación Bill & Melinda Gates, y en colaboración con entidades globales, este proyecto proporciona información actualizada a la Organización Mundial de la Salud (OMS). El carácter vivo de esta revisión presentó diversos desafíos metodológicos. Este proyecto incluye varias etapas. En primer lugar, se diseñó un formulario en RedCap (sistema de captura de datos electrónico) para recopilar los datos de la revisión sistemática, adaptándose a la diversidad de información de los artículos recolectados. Luego, se creó la plataforma web que integra: (1) un tablero desarrollado en Power BI que permite visualizar los resultados actualizados de la revisión, y (2) una aplicación web desarrollada en Shiny de R que muestra los resultados de diversos meta-análisis de interés. Ambas herramientas se actualizan automáticamente con los datos de RedCap. Los artículos científicos incluidos en la revisión reportan información para diferentes outcomes (resultados) y diferentes subgrupos de interés (plataforma de la vacuna, trimestre embarazo, variante de COVID 19, etc.), permitiendo realizar múltiples meta-análisis considerando todas las posibles combinaciones de estos factores. La aplicación en Shiny facilita la visualización de los resultados de estos meta-análisis seleccionando filtros según outcome y subgrupos. Un reto metodológico importante es evitar (de forma automática) el doble conteo de instancias, ya que los artículos no siempre son homogéneos en la forma de reportar los resultados y los subgrupos pueden no ser independientes. Los hallazgos actualizados están disponibles en <https://www.safeinpregnancy.org/living-systematic-review/>. Esta plataforma es útil para los tomadores de decisiones en salud, proporcionando acceso a evidencia relevante sobre los beneficios y riesgos de las vacunas COVID-19 en personas embarazadas.

Palabras clave: **revisión sistemática, meta análisis, covid-19, vacunas, shiny app.**



Métodos Bayesianos

EFFECTO DE LA MODELACIÓN DE LA PRESENCIA DEL CARBÓN DE LA ESPIGA DEL MAÍZ (*Ustilago maydis*) EN LA IDENTIFICACIÓN DE RESISTENCIA

CECILIA BRUNO¹, ANDREA PEÑAS BALLESTEROS^{2,3}, JULIANA IGLESIAS²

¹*Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba – Grupo de Estadística y Biometría vinculado a la Unidad de Fitopatología y Modelización Agrícola (UFyMA) – INTA-CONICET*

²*Instituto Nacional de Tecnología Agropecuaria- EEA Pergamino. Universidad Nacional del Noroeste de la Provincia de Buenos Aires (UNNOBA)*

³*CITNOBA (UNNOBA - UNSAdA – CONICET), Buenos Aires*

cebruno@agro.unc.edu.ar

RESUMEN

Ustilago maydis es un hongo basidiomicete biótrofo causal del carbón común del maíz (*Zea mays* L.) que se manifiesta como agallas que afectan el rendimiento de los granos y pueden facilitar la infección por otros patógenos. El efecto del patógeno sobre la planta suele registrarse a través de un conteo de cantidad de plantas enfermas sobre número de plantas sanas de una parcela (variable discreta). El mapeo de asociación de genoma completo (GWAS, del inglés *Genome Wide Association Study*) es una herramienta eficaz para identificar loci de caracteres cuantitativos (QTL) a través de la asociación genotipo (SNP)-fenotipo, sin embargo, se espera que la distribución de la variable respuesta (presencia/ausencia de carbón de la espiga de maíz) tenga una distribución normal. En este trabajo se evaluaron ocho modelos lineales generalizados mixtos de tipo inflados en cero. A partir de la estimación de los BLUP de cada genotipo se ajustaron modelos GWAS multilocus para la identificación de asociaciones al carbón de la espiga en maíz. Se usaron modelos lineales generales (GLM) con estructura genética poblacional obtenida por componentes principales (P), matriz de parentesco por el método bayesiano (Q) y kinship por el método de Van Raden (K). También los basados en modelos lineales mixtos (MLM) como Compressed (CMLM), Multiple loci (MLMM), FarmCPU y Blink de GAPIT. Se trabajó con una base de datos de 63 líneas endocriadas de maíz genotipadas mediante un chip de SNPs de 56K proveniente del Programa de Mejoramiento de maíz de la Estación Experimental Agropecuaria INTA Pergamino. El modelo mezcla con la distribución Binomial Negativa (ZINB) fue el que obtuvo menor valor de AIC y los modelos MLMM y CMLM identificaron siete segmentos cromosómicos candidatos para resistencia a carbón.

Palabras clave: **modelos lineales generalizados mixtos, variables discretas de conteo, modelos lineales mixtos, métodos bayesianos, estructura genética poblacional.**

Ciudad de Buenos Aires, octubre 2024

VEROSIMILITUD PENALIZADA CON APORTES BAYESIANOS PARA ANALIZAR LA LOCALIZACIÓN DE TERMORRECEPTORES EN LAS ANTENAS DE *RHODNIUS PROLIXUS*

CHIALINA TOMÁS MANUEL^{1,2}, MINOLI SEBASTIÁN¹, FRASCA CECILIA^{1,3}

¹Laboratorio de Fisiología de Insectos, Departamento de Biodiversidad y Biología Experimental, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires / Instituto de Biodiversidad y Biología Experimental y Aplicada, CONICET-UBA

²Laboratorio de Fisiología de la Visión, Departamento de Fisiología y Biología Molecular y Celular, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires / Instituto de Biociencias, Biotecnología y Biología Traslacional, UBA

³Laboratorio de Investigaciones en Hormigas y otros Artrópodos, Instituto de Investigaciones en Biodiversidad y Medioambiente, CONICET - Universidad Nacional del Comahue. San Carlos de Bariloche

tmanuelch@gmail.com

RESUMEN

El calor es probablemente la señal más relevante que los animales hematófagos utilizan para encontrar alimento. En los insectos triatominos (vinchucas), la respuesta de extensión de la probóscide (REP) para perforar la piel de un hospedador es un proceso mediado por el calor, entre otras claves sensoriales. Aquí, estudiamos la localización intra-antenal de los termorreceptores involucrados en la REP de la vinchuca *Rhodnius prolixus*. Para ello realizamos antenectomías parciales, en las cuales se removieron diferentes artejos de las antenas. En cada animal, luego de 2 horas post-antennectomía, se registró la ocurrencia de REP ante un estímulo térmico en 5 presentaciones separadas por intervalos de 1 minuto. Se realizaron 20 réplicas por tratamiento. Los resultados se analizaron con un modelo lineal generalizado mixto (GLMM) asumiendo una distribución binomial. Debido a la separación completa detectada en nuestros datos—causada por tratamientos con total ausencia de respuestas o bien con total presencia—optamos por utilizar una regresión logística penalizada alternativa al algoritmo estándar de reducción de sesgo propuesto por Firth (1993) para solucionar dicho fenómeno. Esta variante nos permite incorporar factores aleatorios en el modelo, y consiste en la incorporación bayesiana de una distribución *a priori* normal y con media de cero para los efectos fijos en el marco de un GLMM. Además, es fácilmente aplicable en R mediante la función *bglmer* del paquete *blme*. Modelando los datos de esta manera, observamos que al cortar el flagelo más distal de la antena (F2), la probabilidad de REP disminuyó significativamente al 0.25, mientras que al cortar el siguiente flagelo (F1), la REP desapareció por completo. En aquellos animales intactos o con una antena removida, la probabilidad de REP fue de 1. Concluimos que los termorreceptores necesarios para gatillar la REP se encuentran distribuidos principalmente en los F2 de las antenas de *R. prolixus*, y en menor medida, en los F1.

Palabras clave: *glmm - binomial - separación completa - bayesiano - termopercepción*.

Ciudad de Buenos Aires, octubre 2024

ESTIMACIÓN DE VARIANZA GENÉTICA EN LA RESISTENCIA A INSECTICIDAS

GANDINI LUCIANO M.¹, SANTADINO MARINA v.², FANARA JUAN JOSÉ¹

¹ *Instituto de Ecología, Genética y Evolución (IEGEB), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires-CONICET*

² *Departamento de Tecnología, Universidad Nacional de Luján.*

lgandini@ege.fcen.uba.ar

RESUMEN

Drosophila suzukii es una mosca plaga que causa graves daños económicos a la producción de frutas. Se estimó la mortalidad de moscas expuestas a tres insecticidas (Lambdacialotrina, Imidacloprid y Spinosad) y la varianza genética de este carácter. Se utilizaron 24 líneas endocriadas para estimar la varianza genética en la población, que puede ser medida como la varianza fenotípica (observada) entre líneas isogénicas. Grupos de cinco moscas, machos y hembras por separado, fueron expuestos a un insecticida en dos concentraciones y, se determinó el número de moscas muertas luego de 12, 24 y 36 horas. Los análisis estadísticos se desarrollaron en R con el paquete brms, que implementa un algoritmo bayesiano. Se modeló la mortalidad a tiempos discretos como un GLMM de distribución beta-binomial con los factores fijos tipo de insecticida, sexo, tiempo, concentración y sus interacciones. Los factores aleatorios fueron la réplica, un efecto aleatorio a nivel de la observación, la línea isogénica (variación entre líneas en la mortalidad, independiente del tratamiento) y las interacciones línea-sexo, línea-insecticida y línea-concentración (variación expresada diferencialmente ante cada tratamiento). Se calculó el coeficiente de correlación intraclase (CCI) para cada factor aleatorio. El modelo fue validado a través de verificaciones predictivas posteriores (“pp-checks”). El efecto de los insecticidas es dependiente de la dosis y el tiempo desde la aplicación. Lambdacialotrina y Spinosad presentan mayor mortalidad con la dosis más alta pero sus máximos son a las 12 y 24 h, respectivamente. La mortalidad con Imidacloprid es más baja en todos los tiempos y concentraciones. El CCI del factor línea representa un 8,38% de la variación total, la interacción línea-sexo un 10,08%, línea-concentración un 7,12% y línea-insecticida 17,76%. Gran parte de la variación genética está dada por la interacción línea-insecticida, por lo tanto, la susceptibilidad de cada línea depende del insecticida aplicado.

Palabras clave: ***Insecticidas, resistencia, varianza genética, componentes de varianza.***

ESTIMACIÓN DE LA EDAD DE LOS ROEDORES USANDO MODELOS BAYESIANOS

JULIETA PELLETTIERI¹, MARIANO MARZIALI BERMÚDEZ², IRENE L. GOROSITO³

¹*Grupo de Bioestadística Aplicada, Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

²*Departamento de Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Instituto de Física de Buenos Aires - CONICET*

³*Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Instituto de Ecología, Genética y Evolución de Buenos Aires - CONICET*

ireneg@ege.fcen.uba.ar

RESUMEN

Los métodos de estimación de la edad de roedores que requieren sacrificar a los animales no pueden aplicarse en estudios de seguimiento poblacional de individuos. Se propone un método de estimación de edades de roedores a partir de medidas del largo del cuerpo y del peso, basado en una versión modificada del modelo de crecimiento de Richards, donde cada rasgo (y = peso o largo) sigue una relación con la edad ($x = t - b$, siendo t la fecha de observación y b , la de nacimiento) de la forma $y(x) = a_0 + a_1(1 - e^{-x/T} + r x)$. El método se aplicó a datos de 1314 individuos de la especie *Akodon azarae*, recolectados durante 11 muestreos de captura y recaptura de roedores en un ambiente rural, de los cuales 247 fueron capturados en más de un muestreo. Los parámetros del modelo, incluyendo la fecha de nacimiento, de cada individuo, se estimaron mediante la técnica bayesiana de muestreo de Gibbs, utilizando el paquete jags para R. Todos los parámetros del modelo se trataron de forma jerárquica, permitiendo variabilidad entre individuos (curvas de crecimiento individuales), de acuerdo con priors determinados sobre la base de un primer ajuste a datos de bioterio. Las medias de las edades resultaron de entre 13 días y 30 meses, y sus distribuciones posteriores exhibieron desvíos estándar, mayormente, de entre 2 y 3 meses, observándose mayor dispersión en los individuos de mayor edad (desde 5 días para los más jóvenes y hasta 6.7 meses para los roedores más viejos). Esto resulta esperable, ya que cuando el crecimiento es más lento, la variabilidad entre individuos se vuelve predominante por sobre la variación promedio debido al crecimiento. Más allá de estas limitaciones, este método evita la necesidad de dividir la población por grupos discretos, permitiendo conocer la estructura de edades de una población de roedores con una cota de credibilidad. Además, bajo supuestos de clausura, se pueden realizar estimaciones de abundancia más precisas cuando la detectabilidad no es perfecta, extendiendo el alcance del mínimo número vivo conocido hasta el momento del nacimiento.

Palabras clave: ***inferencia bayesiana, estimación de edad, modelos de crecimiento.***

Ciudad de Buenos Aires, octubre 2024



Métodos Estadísticos

CÁLCULO DE UNA f DE COHEN EFICIENTE PARA UNA PRUEBA ÓMNIBUS

MORALES, ANDRÉS ¹

¹ Estación Experimental Agropecuaria La Consulta, Instituto Nacional de Tecnología Agropecuaria. Cátedra de Genética General y Aplicada, Facultad de Ciencias Agrarias, Universidad Nacional de Cuyo. Laboratorio de Biología Molecular, Instituto de Biología Agrícola de Mendoza (IBAM – CONICET).

morales.andres@inta.gob.ar

RESUMEN

El tamaño muestral para un correcto diseño experimental depende de la magnitud del efecto a detectar. En el caso de un factor con dos niveles ($k = 2$), la d de Cohen mide este efecto utilizando la diferencia de medias, mientras que para $k \geq 2$, la f de Cohen mide el desvío entre medias, en ambos casos, relativos al desvío dentro de grupos. Para diseñar un ensayo, entonces, resulta intuitivo definir la magnitud del mínimo efecto a detectar en términos de d y calcular el tamaño muestral a partir de su correspondiente en la escala de f . Sin embargo, la equivalencia entre d y f no es unívoca: depende de la distribución de las k medias de grupos. Cohen (1988) plantea fórmulas de f en función de d para tres escenarios: i) varianza mínima (un grupo con efecto $d/2$, otro $-d/2$ y el resto 0), ii) intermedia (medias distribuidas equidistantes en $[0, d]$) y iii) máxima ($k/2$ ó $(k+1)/2$ grupos con efecto d y el resto 0). Estos escenarios, no obstante, son casos particulares de la hipótesis alternativa de la prueba ómnibus del análisis de la varianza. Aquí, presento una fórmula alternativa para alcanzar la potencia requerida para detectar al menos un efecto d . Para verificarlo, calculé la potencia derivada de estas fórmulas en 10^4 simulaciones para diversas combinaciones de k y d . La potencia estadística lograda se mantuvo en los valores predefinidos ($p = 0.593$). La fórmula alternativa presentada para el cálculo de f es más eficiente, es decir, requiere un menor tamaño muestral que la fórmula i) para $k > 2$ y tiene más potencia que iii) para $k > 3$. Con respecto a la fórmula ii), en el intervalo $[3 < k < 8]$ la fórmula alternativa es más eficiente y cuando $k > 8$, más potente.

Palabras clave: ***magnitud de efecto, f de cohen, análisis de la varianza***



Métodos Robustos y No Paramétricos

HERRAMIENTAS PARA EL ANÁLISIS ROBUSTO DE DATOS DOSIS-RESPUESTA

A. M. BIANCO¹, G. CHAUFAN², I. COALOVA², C. PITA¹, M. VALDORA^{1,3}

¹*Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires—
CONICET*

²*Dpto. de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires—
IQUIBICEN-CONICET*

³*Dpto. de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires—
CONICET*

carlosjosepita@gmail.com

RESUMEN

Describimos la implementación de un paquete en R para la estimación robusta y eficiente de posición en modelos de dosis-respuesta. Se trata de un M-estimador redescendente con una estructura de tres pasos, semejante a la de los MM-estimadores: primero se calcula un estimador robusto inicial de posición, con punto de ruptura alto; luego un estimador de escala también robusto y que contempla heterocedasticidad; finalmente, iterando a partir de los valores iniciales, el M-estimador deseado que combina robustez y eficiencia. La implementación soporta modelos no lineales, heterocedásticos y con presencia de outliers. Ofrece una paleta de estimadores robustos configurables para cada paso y la posibilidad de indicar otros estimadores y modelos dosis-respuesta paramétricos externos. Como ejemplo de aplicación, presentamos la estimación —tanto sobre datos reales como simulados— de la condición aditiva, sinérgica o antagónica de la citotoxicidad de dos agentes químicos en células humanas, empleando un modelo logístico de 4 parámetros (4PL) y la combinación de un S-estimador bisquare para el paso 1, un M-estimador de escala bisquare para el paso 2 y un M-estimador bisquare para el paso 3. Para concluir, comentamos sobre algunos desafíos computacionales y estadísticos que derivan de las desviaciones de los supuestos típicos de linealidad y homocedasticidad, así como del empleo de estimadores redescendentes.

Palabras clave: **dosis-respuesta, robusto, heterocedástico, no-lineal, R, paquete**



INTERPRETANDO LAS DESVIACIONES DE LA PREDICCIÓN ALOMÉTRICA EN MACROEVOLUCIÓN: UN ENFOQUE UTILIZANDO TASAS DE CAMBIO

IGLESIAS PP¹, CASTIGLIONE S², ELIAS-COSTA A^{3,4}, BALDO D⁵, PEREYRA MO¹

¹ CONICET-Agencia INTA General Acha, Estación Experimental Anguil, General Acha, La Pampa

² Department of Earth Sciences, Environment and Resources, University of Naples Federico II, Naples, Italy

³ Herpetology, Museum für Naturkunde, Leibniz Institute for Research on Evolution and Biodiversity Berlin, Germany

⁴ División Herpetología, Museo Argentino de Ciencias Naturales "Bernardino Rivadavia"-CONICET, Buenos Aires

⁵ Instituto de Biología Subtropical (IBS, CONICET), Universidad Nacional de Misiones

patricia.p.iglesias@gmail.com

RESUMEN

La variación en la magnitud relativa de un carácter respecto del tamaño corporal (TC) se interpreta comúnmente como variación independiente del TC. Sin embargo, esto podría ignorar la influencia del TC en la relación alométrica. Las relaciones alométricas revelan patrones específicos de coevolución, pero no logran revelar qué procesos subyacen a la desviación residual de las especies. En este estudio utilizamos un enfoque que considera tasas de cambio para entender mejor estos procesos, y lo aplicamos al estudio de la alometría acústica en anuros. Para ello: 1) modelamos la evolución de la frecuencia dominante (FD) y el TC independientemente para obtener tasas de cambio absolutas y reconstruir estados ancestrales mediante regresión Ridge; 2) utilizamos permutaciones para identificar cambios significativos en las tasas absolutas; 3) estimamos cambios proporcionales basados en los estados ancestrales; y 4) utilizamos regresiones de tipo II (Ranged Major Axis, RMA) para identificar los sesgos en las desviaciones del patrón alométrico en cada grupo identificado en el punto 2. Esta aproximación se basa en que la pendiente alométrica describe el cambio proporcional desde el ancestro hasta el descendiente; al graficar tasas de cambio en lugar de los valores fenotípicos, se produce un colapso de la línea alométrica en isometría. Así, la línea isométrica representa la relación alométrica entre FD-TC, independiente del valor del coeficiente alométrico. Encontramos que el cambio proporcional en la FD es siempre mayor o igual que el cambio proporcional en el TC, independientemente de los procesos que operan acelerando o desacelerando las tasas absolutas. Por lo tanto, la suposición de que la FD relativa refleja principalmente la variación en FD es válida para la relación FD-TC en Anura con excepción del clado Dendrobatidae. Desde una perspectiva perceptual, la FD se identifica como un parámetro fácil de discriminar y adaptar a entornos cambiantes.

Palabras clave: **Alometría Acústica, Anuros, Permutaciones, Regresión Ridge, Regresión tipo**

Ciudad de Buenos Aires, octubre 2024



Modelos Aditivos Generalizados

LA EVAPOTRANSPIRACIÓN EN BOSQUES MISIONEROS CON Y SIN EXTRACCIÓN SELECTIVA DE MADERA

PARONETTO JS^{1,2}, DIAZ VILLA MVE^{1,2}, GOLDSTEIN GH¹ Y CRISTIANO PM^{1,2}

¹ *Laboratorio de Ecología Funcional, Departamento de Ecología Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

² *Instituto de Ecología, Genética y Evolución (IEGEB), Universidad de Buenos Aires-CONICET*

julietaparonetto@gmail.com

RESUMEN

Debido a la gran pérdida de la cobertura boscosa y a la alteración de los bosques remanentes por actividades forestales, resulta prioritario comprender cómo se modifica la evapotranspiración (ET), un proceso clave del ciclo hidrológico. El objetivo de este trabajo fue analizar y comparar la ET de bosques subtropicales húmedos de la provincia de Misiones: bosques conservados (BC) y bosques con extracción selectiva de madera (BE), con distintos tiempos desde la última extracción (BE>30, 10-15, 5-10, 1-5, 0-1 años). Se extrajeron datos de ET del producto MOD16A2 del sensor MODIS (*Moderate Resolution Imaging Spectroradiometer*) coincidentes con las parcelas a campo, con resolución espacial de 500m y resolución temporal de 8 días que se escalaron a datos mensuales, desde 2001 hasta 2023. Se ajustó un modelo *gam* (Modelos Generalizados Aditivos) del paquete *mgcv* con una familia *Tweedie* dada la falta de normalidad en la distribución de la variable ET. Se modeló la ET en función del tipo de bosque con dos suavizados: uno para capturar la variación intraanual con interacción entre mes y año, y otro para capturar la variación interanual para cada tipo de bosque. Ambos suavizados fueron significativos y en particular, la variación interanual para cada tipo de bosque fue significativa. Se encontró que hay diferencias en la ET entre distintos bosques ($p < 0.05$). Sin embargo, la ET de los BE no difirió respecto a la de BC. La diferencia se observó entre la ET de los BE 5-10 años que fue significativamente mayor a la de los BE 10-15 y 0-1 años. Estos resultados sugieren que la ET de los bosques misioneros no se modificaría por la extracción selectiva de madera. Las diferencias en la ET de los BE podrían estar relacionadas con la estructura de la vegetación.

Palabras clave: **modelo GAM, series temporales, teledetección**

DESIGUALDADES EN LA REALIZACIÓN DE ACTIVIDAD FÍSICA Y DE SEDENTARISMO EN ADULTOS SEGÚN CARACTERÍSTICAS CLIMÁTICAS Y SOCIODEMOGRÁFICAS EN ARGENTINA: UN ESTUDIO A PARTIR DE LA ENCUESTA NACIONAL DE FACTORES DE RIESGO 2018

VALENTIN ROZENBLIT¹, GABRIELA DE ROIA², GERARDO CUETO¹, ADRIANA PÉREZ¹

¹ Grupo de Bioestadística Aplicada, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Buenos Aires

² Laboratorio de Estudios en Actividad Física (LEAF), Universidad de Flores, Buenos Aires
valentin.rozenblit@gmail.com

RESUMEN

La baja actividad física (AF) y el comportamiento sedentario (CS) son factores de riesgo para enfermedades no transmisibles. Globalmente, enfrentamos desafíos para reducir la prevalencia de AF baja y CS debido a barreras como las condiciones climáticas y sospechamos que el efecto de éstas cambia según el nivel socioeconómico (NSE) de los individuos. En este estudio, investigamos la relación entre la AF baja y CS, el clima y el NSE. Se trabajó con la Encuesta Nacional de Factores de Riesgo 2018, conformada por una muestra con cobertura nacional de 29224 individuos mayores de 18 años y con datos climáticos obtenidos del repositorio mundial WorldClim. El NSE a escala individual se obtuvo realizando un Análisis de Clases Latentes utilizando como variables indicadoras del NSE al quintil de ingresos, nivel de instrucción, carencias materiales y de vivienda. Se implementaron modelos aditivos generalizados (GAM) y regresiones logísticas para analizar la relación entre la realización de AF baja, variables climáticas a escala provincial (temperatura media anual (TMA) y precipitación media anual (PMA)) y NSE a escala individual, controlando por género. Se encontró una asociación no lineal entre AF baja y TMA, con una mayor proporción de personas con AF baja a TMA por debajo de los 17 °C y por encima de los 19 °C, y una asociación inversa con NSE. Por otro lado, se encontró una asociación lineal directa entre AF baja y PMA, con una interacción significativa con NSE. A mayor PMA, mayor proporción de personas con AF baja, siendo más fuerte esta asociación en personas con bajo NSE. Y en cuanto al CS, se encontró una relación lineal indirecta con TMA y PMA, con una interacción significativa con NSE. A mayor TMA y PMA, menor CS solo para los grupos de menor NSE. Se concluye que las desigualdades en ciertas variables climáticas amplifican las desigualdades en mantener una vida activa y supone la necesidad de mejorar el acceso a oportunidades asequibles de AF, que no se vean afectadas por el clima junto con una mejor concientización sobre el CS.

Palabras clave: **modelos aditivos generalizados, regresión logística, datos públicos, factores de riesgo, insuficiente actividad física, factores climáticos.**

Ciudad de Buenos Aires, octubre 2024



Modelos de Simulación

MODELO ESTOCÁSTICO DE LA DINÁMICA DE PROPAGACIÓN DE PESTE PORCINA AFRICANA EN EL SECTOR PORCINO DE LA ARGENTINA

ALARCON, L.V.¹; LA SALA, L.F.²; PEREZ, A.³; CIPRIOTTI, P.A.⁴

¹ *Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, La Plata*

² *Instituto de Ciencias Biológicas y Biomédicas del Sur (INBIOSUR), CONICET – Universidad Nacional del Sur, Bahía Blanca.*

³ *Servicio Nacional de Sanidad y Calidad Agroalimentaria (SENASA), Ministerio de Producción y Trabajo, Buenos Aires.*

⁴ *IFEVA, Facultad de Agronomía, Universidad de Buenos Aires- CONICET*

lalarcon@fcv.unlp.edu.ar

RESUMEN

Se construyó un modelo de simulación mecanicista, espacialmente explícito y estocástico basado en las condiciones argentinas, con un paso temporal diario y una representación espacial real de las entidades implicadas en la propagación de la Peste Porcina Africana (PPA). Simulamos la propagación en dos niveles: en el superior, la PPA se propaga entre granjas u entidades espaciales a través de vías directas (movimiento de animales y vehículos entre nodos de la red) y vías indirectas (contactos con jabalíes, fómites o residuos de alimentos humanos). En el nivel inferior, una vez infectada una granja, se utilizó un modelo SEIR para simular la dinámica de PPA intragranja. Los parámetros del modelo se calibraron a partir de datos de campo y fuentes bibliográficas. Luego, se realizó un análisis de sensibilidad para evaluar el papel de cinco vías de transmisión: granjas grandes o pequeñas, poblaciones de jabalíes, llegada de alimentos contaminados a basureros y camiones contaminados. Durante los primeros 50-70 días desde los episodios iniciales de infección, el papel de los camiones y de las granjas pequeñas (10-50 cerdas) mediante el intercambio de personal o utensilios y la adquisición de desperdicios, aumentó la propagación y prevalencia, superando a las demás vías. La extensión espacial de los contactos (por ejemplo, personal y fómites) entre granjas a 10-35 km fue clave en controlar la dinámica. El modelo que incluyó todas las vías, alcanzó una prevalencia a nivel de red del 61% a los 60 días, mientras que una reducción del radio de contactos de 25 km la redujo al 12%. La prevalencia media intra-granja fue del 24,4%, pero con una considerable heterogeneidad espacial en función del tipo de granja. Además, la prevalencia de las poblaciones de jabalíes era de un 2%, mientras que el porcentaje de camiones contaminados era de un 4% del total del país.

Palabras clave: **modelo estocástico, peste porcina africana, transmisión enfermedades, jabalíes.**



Modelos de Supervivencia

DEL EFECTO DEL RIESGO DE DEPREDACIÓN SOBRE LA LATENCIA DE RETORNO AL NIDO MEDIANTE MODELOS DE REGRESIÓN DE COX

DURE RUIZ NATALIA M.¹, FERNÁNDEZ GUSTAVO J.²

¹*Departamento de Biodiversidad y Biología Experimental, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

²*Centro Regional San Martín de los Andes, INIBIOMA-Universidad Nacional del Comahue*

natymdr@ege.fcen.uba.ar

RESUMEN

Los métodos Kaplan-Meier y modelos de Regresión de Cox permiten analizar efectos de variables predictivas sobre el tiempo hasta que ocurre un evento. Este trabajo estudia el efecto del riesgo de depredación percibido por la Ratona Común (*Troglodytes musculus*) sobre el tiempo de retorno al nido. Durante 4 años parejas de Ratona se expusieron a aves que representan diferente riesgo, todas embalsamadas en posición reposo, (3 depredadores simpátricos, 1 novel y 1 no-depredador); 84 parejas durante el período incubación y 97 durante el período cría de pichones. La exposición duró 25 minutos, cada pareja se expuso sólo a un ave por período. Se midió el tiempo transcurrido desde que inicia la exposición hasta que machos y hembras retornaron al nido (latencia, en segundos). Para cada período se construyeron modelos donde se introdujeron dos variables dependientes: latencia y retorno (variable dicotómica 1=retorna/0=no retorna); tres predictoras: riesgo (un nivel por ave presentada), etapa de cuidado (temprana/tardía) e interacción riesgo*etapa; una aleatoria: identidad del nido (control de variabilidad de respuesta). Análisis previos con otro depredador simpátrico, descartaron variaciones en la latencia por efecto año (Kaplan-Meier, $p > 0.05$). Los modelos revelaron variaciones significativas en la latencia por efecto del riesgo de depredación. Machos y hembras retrasaron significativamente el retorno al nido frente una amenaza (Incubación: $X^2=14.56$ GL=2 $p < 0.01$ hembras; Cría: $X^2=32.77$ GL=3 $p < 0.001$ hembras, $X^2=35.53$ GL=3 $p < 0.01$ machos). Comparaciones pareadas de Tukey revelaron mayor retraso a mayor riesgo para la pareja ($p < 0.05$ en todas las comparaciones), no detectaron diferencias entre un depredador de bajo riesgo vs un no-depredador ni entre un depredador novel vs un depredador de similar riesgo ($p > 0.05$ en ambas comparaciones). El estudio revela que la Ratona Común reconoce y discrimina depredadores simpátricos, ajustando la respuesta al grado de amenaza. Además, identifica depredadores no-simpátricos, pero falla al determinar su grado de amenaza.

Palabras clave: **regresión de cox, troglodytes musculus, reconocimiento de depredadores**

SUPERVIVENCIA APARENTE Y TASA DE RETORNO DEL JILGUERO DORADO (*SICALIS FLAVEOLA PELZELNI*), UNA ESPECIE CON FENOTIPO RELACIONADOS CON LA EDAD

PALMERIO ANDRÉS¹, MASSONI VIVIANA¹²

¹*Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires*

²*Instituto de Ecología, Genética y Evolución - IEGEBA, Universidad de Buenos Aires*

apalmerio@ege.fcen.uba.ar

RESUMEN

Estudiamos la tasa de retorno, la supervivencia aparente y la probabilidad de recaptura del Jilguero dorado (*Sicalis flaveola pelzelni*), un tráupido con dicromatismo sexual y fenotipos del plumaje masculino relacionados con la edad; los machos del segundo año son opacos e indistinguibles de las hembras tanto para los humanos como para sus congéneres, mientras que los machos mayores de dos años son de color amarillo dorado. El estudio abarcó seis temporadas reproductivas. Se utilizaron modelos para datos de encuentros vivos, corregidos por la probabilidad de encuentro (P) para analizar la supervivencia de los adultos. Las tasas de supervivencia aparente (ϕ) y recaptura (P) de los adultos se estimaron utilizando modelos de captura-recaptura para poblaciones abiertas de Cormack-Jolly-Seber (CJS) con el Programa MARK 9.0 utilizando criterios de máxima verosimilitud. Hallamos que los machos y hembras con plumaje adulto no difirieron en la tasa de retorno, pero los machos con plumaje inmaduro tuvieron tasas de retorno más bajas que los machos con plumaje adulto. Las tasas de retorno no variaron con el éxito de nidificación previo. La tasa de supervivencia aparente de todos los machos combinados y de las hembras fue casi igual, pero encontramos diferencias importantes cuando comparamos ambos tipos de machos, ya que los machos mayores tuvieron una tasa de supervivencia aparente más alta que los más jóvenes. La probabilidad de recaptura estuvo relacionada con el sexo y la edad, teniendo las hembras una mayor probabilidad de recaptura que los machos, y los machos mayores una probabilidad de recaptura más alta que los machos jóvenes. Finalmente, comparamos la supervivencia aparente de Jilgueros dorados con aquellas de estudios publicados sobre tráupidos y resultó ser similar al promedio de la familia.

Palabras clave: *supervivencia aparente; maduración tardía del plumaje; tasa de retorno; Sicalis flaveol; tráupidos*

ECTOPARASITICIDAS APLICADOS A GATOS DOMÉSTICOS COMO HERRAMIENTA DE CONTROL DE *T. INFESTANS* RESISTENTES A LOS PIRETROIDES

CAMILA VAZQUEZ-CAÑAS¹, MARIA S. GASPE¹, GUSTAVO F. ENRIQUEZ¹, DAMIAN SANCHEZ, DELFINA TREZZA-NEUMAYER¹, BARBARA L. OJEDA¹, SANTIAGO PIÑERO¹, MARTA V. CARDINAL¹

¹Laboratorio de Eco-Epidemiología, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires- IEGEBA-CONICET

cami.vazqc@gmail.com

RESUMEN

El Gran Chaco es una región hiperendémica para la enfermedad de Chagas, con elevados niveles de infestación por *Triatoma infestans* y transmisión vectorial de *Trypanosoma cruzi*. Con la emergencia de poblaciones de *T. infestans* resistentes a insecticidas piretroides se vuelve crucial incorporar herramientas de control alternativas, como la xenointoxicación. El objetivo de este estudio es determinar la eficacia triatomicida y duración de distintos ectoparasiticidas aplicados en gatos domésticos del Municipio de Pampa del Indio, Chaco. Se seleccionaron 25 gatos de viviendas rurales mediante un diseño aleatorizado por grupos. Se emplearon 3 ectoparasiticidas (fluralaner, spinosad e imidacloprid) y 2 controles (agua y comida) acordes a las formulaciones empleadas (spot-on y comprimido), aplicados en una sola dosis. Se expusieron *T. infestans* (III-V estadio) (n=1102) susceptibles (S) y resistentes (R) a piretroides, a los 0, 4, 30 y 60 días post-tratamiento (DPT). Se registró la mortalidad a 0-5 días y luego dos veces por semana durante 4 semanas. Se construyeron curvas de supervivencia de Kaplan-Meier para cada tratamiento, comparadas mediante la prueba de log-rank pareada. Se hallaron diferencias significativas entre fluralaner ($p < .0001$) y su control para los tiempos 4 a 60 DPT y entre spinosad y su control para los tiempos 4 y 30 DPT ($p < .05$), no así para imidacloprid. Utilizando modelos lineales generalizados mixtos obtuvimos una interacción significativa entre el tiempo y el tratamiento con fluralaner y spinosad sobre la mortalidad acumulada a los 20 días de *T. infestans* y no significativa para el tipo R o S. La probabilidad de morir a los 4 y 30 DPT fue 45 y 13 veces mayor respectivamente, respecto al control para fluralaner ($p < .001$) y de 14 y 5 veces mayor para spinosad ($p < .05$). Estos resultados preliminares respaldan el uso de fluralaner y medida, para la xenointoxicación usando gatos domésticos para el control de triatominos resistentes a los piretroides.

Palabras clave: **Xenointoxicación - *T. infestans* - Gran Chaco - resistencia a los piretroides**



Modelos Lineales Generales

ANÁLISIS DE VARIABLES NORMALES DISCRETIZADAS: UN CASO DE ESTUDIO

MORALES, ANDRÉS¹

¹ Estación Experimental Agropecuaria La Consulta, Instituto Nacional de Tecnología Agropecuaria. Cátedra de Genética General y Aplicada, Facultad de Ciencias Agrarias, Universidad Nacional de Cuyo. Laboratorio de Biología Molecular, Instituto de Biología Agrícola de Mendoza (IBAM – CONICET)

morales.andres@inta.gob.ar

RESUMEN

La discretización es un proceso ampliamente utilizado en la clasificación de productos agrícolas en categorías mutuamente excluyentes en base al peso, diámetro, longitud o una combinación de los mismos. Dada la naturaleza numérica discreta de estos datos, los residuos de un modelo lineal presentan patrones “en escalera” que no se ajustan a una distribución normal, y por lo tanto, es práctica común recomendar el uso de métodos estadísticos no paramétricos. No obstante, se puede argumentar que en caso de que la variable original se distribuya normalmente, los resultados de un modelo lineal sobre la variable discretizada son válidos. Para poner a prueba esta hipótesis, en este trabajo analizo las consecuencias de la discretización de una variable normal utilizando el caso de un ensayo de dos planes de fertilización de ajo en un diseño completamente aleatorizado de 5 parcelas con 8 submuestras ($n = 80$). El diámetro mayor de bulbos se midió utilizando un calibre manual y se ajustó un modelo lineal. La variable original se redondeó y la variable discretizada se analizó utilizando un modelo lineal y dos métodos no paramétricos. Con el fin de realizar inferencias, se realizaron 10^4 remuestreos con reposición. Se observó una pequeña reducción en la media de la variable discretizada con respecto a la original ($-0,368\%$, $p < 2 \cdot 10^{-16}$) y un gran aumento en la varianza ($39,88\%$, $p < 2 \cdot 10^{-16}$), lo que conlleva una disminución en la d de Cohen de $16,8\%$ y una pérdida de potencia observada en el modelo lineal de $3,69\%$ al discretizar la variable. Al comparar este último modelo con los no paramétricos se observa una significativa pérdida de potencia de $5,26\%$ y $22,2\%$ usando los tests de Kruskal–Wallis y Chi Cuadrado respectivamente. En conclusión, para tamaños muestrales pequeños se recomienda ajustar modelos paramétricos utilizando la variable discretizada.

Palabras clave: **discretización, datos discretos, análisis de la varianza**

MODELO FACTORIAL DE ANOVA PARA LA BÚSQUEDA DE ANTIFÚNGICOS AMIGABLES CON EL MEDIO AMBIENTE Y SU APLICACIÓN EN EL PATRIMONIO CULTURAL

CAMILA S. VOVCHUK^{1,2}, TOMÁS GONZÁLEZ GARELLO³, ALEJANDRA T. FAZIO², VALERIA P. CAREAGA¹

¹UMYMFOR (UBA-CONICET), Departamento de Química Orgánica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

²INMIBO (UBA-CONICET), Departamento de Biodiversidad y Biología Experimental, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

³Grupo de Bioestadística Aplicada, Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

camilavovchuk@gmail.com; tgarello@fmed.uba.ar

RESUMEN

Los bienes realizados en madera que conforman el patrimonio cultural presentan dificultades frente a condiciones ambientales desfavorables y al ataque de microorganismos causantes de biodeterioro. Dentro del diverso grupo de microorganismos, los hongos xilófagos son los principales descomponedores de madera, por lo que su capacidad degradativa puede comprometer la integridad de los objetos realizados en este material. A su vez algunas maderas presentan extractos naturales que le sirven como fuente de protección ante diversos microorganismos. En este trabajo evaluamos la potencial actividad antifúngica de extractos obtenidos con cuatro solventes (metanol, acetona, acetato de etilo y diclorometano) de tres maderas: Eucalyptus camaldulensis, Cedrela fissilis y una madera del género Pinus, sobre tres hongos xilófagos: Chaetomium globosum (Ch.g.), Pycnoporus sanguineus (P.s.), Trametes trogii (T.t.). Para cada ensayo, evaluamos dos concentraciones del extracto (0,3 mg/ml y 0,5 mg/ml). Realizamos tres réplicas de cada tratamiento. Cada hongo se incubó junto al correspondiente y luego de 7 días se calculó el porcentaje de inhibición. Modelamos un ANOVA para evaluar qué variables tuvieron efecto en el porcentaje de inhibición, considerando el tipo de madera, la concentración, el solvente y el tipo del hongo; Por conocimientos previos, incluimos en el modelo un término de interacción entre Hongo y Solvente. Para determinar qué grupos difieren entre sí, realizamos pruebas post hoc de Tukey usando un nivel de significancia de 0.05. Así, respecto al porcentaje de inhibición, concluimos que la concentraciones de extracto difieren significativamente, que los 3 tipos de madera difieren significativamente, y que hay diferencias significativas entre solventes, excepto para el hongo Ch.g. La importancia de estos resultados radica en el potencial uso de los extractos de origen natural como antifúngicos, para la preservación de la madera en objetos del patrimonio cultural.

Palabras clave: **extractivos, biodeterioro, patrimonio cultural**

VIRULENCIA PARASITARIA EN RATAS COINFECTADAS: COMPARACIÓN DE DOS MÉTODOS ANALÍTICOS

SANDRA P. ROBALINO NIETO¹, PABLO F. CUERVO^{2,3}, PABLO M. BELDOMENICO¹, ANDREA L. RACCA¹, CINTIA C. PALAVECINO⁴

¹Laboratorio de Ecología de Enfermedades, Instituto de Ciencias Veterinarias del Litoral (ICiVet Litoral), Universidad Nacional del Litoral -CONICET, Esperanza, Santa Fe.

² Departamento de Parasitología, Facultad de Farmacia, Universidad de Valencia, Valencia, España

³ CIBER de Enfermedades Infecciosas, Instituto de Salud Carlos III, Madrid, España

⁴Centro de Estudios de Variabilidad y Cambio Climático (CEVARCAM), Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional del Litoral

ccpala@gmail.com

RESUMEN

La virulencia es el grado de reducción en la salud que experimenta el hospedador ante una infección, lo cual depende tanto del hospedador como del parásito. El objetivo es evaluar virulencia frente a *Trichinella spiralis* y/o *Trypanosoma cruzi* en ratas de laboratorio (*Rattus norvegicus*, var. Wistar/Cmedc), a partir de distintos métodos analíticos. Se realizó un experimento con 32 ratas macho adultas, que fueron inoculadas con uno o ambos parásitos. Seis semanas post-infección se determinó la carga de ambos parásitos y los niveles plasmáticos de las enzimas alanina aminotransferasa (ALT) y aspartato aminotransferasa (AST). Para estimar la virulencia, se empleó un método de evaluación cuantitativa, basado en la reducción de la condición de salud a una carga parasitaria media; y un método de evaluación cualitativa, basado en la reducción de la condición de salud en animales infectados comparados con no infectados. En ambos casos, se aplicaron modelos lineales mixtos con el software R, donde el nivel de cada enzima fue la variable respuesta. En el método de evaluación cuantitativa las variables predictoras fueron la carga parasitaria y la ocurrencia de uno u otro parásito. A partir de los coeficientes de dicho modelo, se calculó la reducción en la condición de salud para una carga parasitaria media. Para el método cualitativo, las variables predictoras fueron las ocurrencias de ambos parásitos. En el método cuantitativo la reducción de la condición de salud fue menor al 10% en ambos parásitos, mientras que en el método cualitativo se evidenció una mayor virulencia en animales infectados con *T. cruzi*. El método cuantitativo mide el efecto de la cantidad de parásitos, ofreciendo un análisis más detallado. El método cualitativo se enfoca en la presencia o ausencia de la infección, simplificando el análisis. Ambos métodos tienen fortalezas y debilidades, y su elección depende del objetivo del estudio.

Palabras clave: *Trichinella spiralis*, *Trypanosoma cruzi*, modelos lineales mixtos, enzimas hepáticas

Ciudad de Buenos Aires, octubre 2024

EFFECTOS DE LA SOBREALIMENTACIÓN POSTNATAL TEMPRANA EN RATAS ADULTAS, SU MODELADO A TRAVÉS DE UN DISEÑO DE MEDIDAS REPETIDAS

PAMELA FERNÁNDEZ^{1,2}, JOSEFINA CARRIÓ³, CORA STOKER^{1,4}, GUILLERMINA CANESINI^{1,2}, JORGE GUILLERMO RAMOS^{1,4}, STELLA VAIRA³

¹*Instituto de Salud y Ambiente del Litoral (ISAL), CONICET / Universidad Nacional del Litoral*

²*Cátedra de Nutrición en Situaciones Patológicas, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral*

³*Departamento de Matemática y Laboratorio de Investigaciones y Servicios en Bioestadística (LISEB), Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral*

⁴*Departamento de Bioquímica Clínica y Cuantitativa, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral*

svaira@fbc.unl.edu.ar

RESUMEN

En respuesta a la pandemia de obesidad como un problema de salud pública, el modelo de camada reducida (*small litter*, SL), para inducir obesidad temprana y la dieta de cafetería (CAF) que simula la dieta obesogénica actual, son útiles para la investigación. Por ello, se plantea estudiar el efecto de la sobrealimentación neonatal y la exposición a CAF en la adultez, las muestras se componen de ratas macho de la cepa Wistar que fueron criadas en camada reducida (4 crías por madre, SL) o normales (10 crías por madre, NL). Desde el día posnatal (DPN) 21 al 90, fueron alimentadas con *pellet* o dieta de control (CON). Durante 11 semanas recibieron CON o CAF (NL-CON, NL-CAF, SL-CON, SL-CAF; 12±2 ratas/grupo). El modelo contempla dos factores: dieta (efectos fijos: CAF/CON), sobrealimentación (efectos fijos: SL/NL) y el tiempo con 11 mediciones, desde el día 91 hasta el día 167, y como variable respuesta del modelo: el peso en gramos. Se utilizó el paquete *rstatix* (*anova_test*) de R para el análisis. Además de los supuestos de normalidad, detección de outliers y homogeneidad de variancias se analizó el supuesto de esfericidad usando la prueba de Mauchly, que se informa automáticamente cuando se utiliza la función *anova_test()*; además al utilizar *get_anova_table()* se extrajo la tabla ANOVA que se corrigió automáticamente para la eventual desviación de la suposición de esfericidad, la corrección aplicada fue la de Greenhouse-Geisser. El nivel de significancia elegido fue $\alpha = 0,05$. Se pudo verificar el supuesto de normalidad, con Shapiro – Wilks, en todas las combinaciones de los factores (el menor valor p fue de 0,422) y no se encontraron outliers extremos. La dieta CAF produjo un aumento de peso significativo: los valores media \pm DS son CAF – NL = $494 \pm 21,5$ y CAF – SL = $488 \pm 42,5$.

Palabras clave: ***análisis de la variancia, medidas repetidas, dieta y sobrealimentación, diseño experimental.***

RENDIMIENTO Y ESTABILIDAD DE HÍBRIDOS Y VARIEDADES DE COLZA PRIMAVERAL EN MÚLTIPLES AMBIENTES

VIGNEAU F.¹, D. P. RONDANINI², L. E. PUHL³

¹Facultad de Agronomía, Universidad de Buenos Aires

²Departamento de Producción Vegetal, Facultad de Agronomía, Universidad de Buenos Aires

³Departamento de Métodos Cuantitativos y Sistemas de Información, Facultad de Agronomía, Universidad de Buenos Aires

f.vigneau@agro.uba.ar

RESUMEN

Colza (*Brassica napus* L.) es el segundo cultivo oleaginoso del mundo siendo la UE, China y Canadá sus principales productores. En Argentina, su difusión es poca y su adopción lenta, asociado a factores comerciales y a la alta variabilidad en los rendimientos. En los últimos años aumentó la adopción de híbridos modernos de alto rendimiento, pero aparentemente con menor estabilidad que las variedades polinizadas abiertas. El objetivo de este trabajo es evaluar la estabilidad y rendimiento de múltiples genotipos de colza creciendo en ambientes de producción en diferentes regiones de Argentina, identificando las diferencias entre variedades e híbridos. Se utilizaron datos de la red nacional de evaluación de cultivares de colza del INTA, provenientes de 8 localidades con ensayos distribuidos en 10 años (2006 a 2015) que incluyeron 40 genotipos. Debido al recambio de genotipos y localidades en los distintos años la matriz resultante presentó un alto grado de desbalance. Mediante la selección de los ambientes más representativos y la aplicación de un algoritmo basado en la descomposición en valores singulares para estimar los valores faltantes se conformó una matriz completa. Para visualizar y cuantificar los efectos de interacción del genotipo con el ambiente se utilizó el análisis de componentes principales de la matriz de efectos de interacción. Para determinar las diferencias entre híbridos y variedades se utilizó un modelo lineal mixto considerando el tipo de reproducción y el ambiente como efectos fijos y el genotipo como efecto aleatorio. Se identificaron genotipos estables y de alto rendimiento, así como grupos inestables de rendimiento menores para un amplio rango de ambientes. Los híbridos y variedades no presentaron diferencias significativas al 5% en estabilidad ni en rendimiento. Se concluye que existen genotipos de colza de alto rinde y estabilidad adaptados a ambientes locales y que los híbridos no tienen necesariamente menor estabilidad.

Palabras clave: **Brassica napus L., rendimiento, genotipo, ambiente, estabilidad.**



METODO STATICO: LA IMPORTANCIA DEL AGUA DULCE: MONITOREO DE ARROYOS Y RÍOS CORDILLERANOS DE NEUQUEN

ZULIANI R.¹, LAVALLE A.¹, LAFFITTE L.^{2,3}

¹Facultad de Economía, Departamento de Estadística, Universidad Nacional del Comahue, Neuquén

²Dirección General de Biodiversidad y Sostenibilidad, Secretaría de Ambiente, Neuquén

³AUSMA, Facultad de Ciencias Agrarias, Universidad Nacional del Comahue, Neuquén

roxanna.zuliani@faea.uncoma.edu.ar

RESUMEN

El monitoreo y control efectivos de los diferentes cuerpos de agua son una herramienta fundamental para asegurar su sostenibilidad, aportar a la ecología de conservación y el cambio global. Este trabajo utiliza datos relevados en 7 sitios de muestreo de la Cuenca Lacar-Hua Hum contando, por un lado, con una tabla con información de 8 variables ambientales; y otra con frecuencias de 41 familias de macroinvertebrados sobre los mismos sitios. Estas dos tablas se repiten en: 2016, 2019 y 2021 obteniendo una secuencia de tablas ecológicas pareadas. El análisis se realiza con el método estadístico STATICO. El objetivo de esta técnica es obtener la parte estable de la dinámica de las relaciones entre las familias y su entorno. Los resultados muestran que la relación entre las variables y las familias de macroinvertebrados presentan algunas diferencias en estos tres años. La parte estable de estas relaciones consiste principalmente: (1) sobre el primer eje con la mayor parte de la variancia explicada (75%) una combinación con un gradiente que va desde familias menos tolerantes en sitios con aguas menos degradadas a las más tolerantes asociadas a sitios con aguas más contaminadas; (2) sobre el segundo eje, un gradiente determinado por la temperatura. El 2021 se destaca con altas concentraciones de nitrógeno total, fosforo total y fosforo reactivo soluble en los sitios Q1 y Q5 asociados principalmente con las familia Chironomidae y Lumbriculidae. En los tres años, los sitios PP2 y PP3 han mantenido mejores condiciones ambientales con presencia de familias más sensibles, tales como: Gripterygidae y Leptophlebiidae. En cambio, los sitios TO y A5 muestran una disminución en la calidad del agua. Esta investigación pretende brindar nuevas herramientas de análisis de este recurso vital para enfrentar los desafíos actuales y futuros. Todos los cálculos fueron realizados con R.

Palabras clave: **statis, compromiso, interestructura, tablas ecológicas, degradación del agua**



Modelos Lineales Generalizados

EXTRACCIÓN DE HONGOS EN *PEDICULUS HUMANUS CAPITIS*

ATTONATY GABRIELA^{1,2}, CERIANI-NAKAMURAKARE², GONZALES AUDINO PAOLA¹

¹Departamento de Bioseguridad (ex DIPEIN) CONICET-CITEDEF, Villa Martelli, Buenos Aires

²Instituto de Micología y Botánica (INMIBO), CONICET-UBA, Universidad de Buenos Aires

attonaty.gabriela@gmail.com

RESUMEN

Pediculus humanus capitis es un insecto ectoparásito, hematófago que se alimenta de sangre humana. Las infestaciones por este insecto son de distribución mundial, cuyo control representa una gran problemática. Mediante el análisis de la microbiota asociada, se espera poder encontrar algún tipo de control biológico de este insecto. Se realizaron 23 muestreos iniciales donde se recolectaron 237 individuos, 189 hembras y 52 machos. En cada muestreo se realizó el siguiente diseño experimental: se recolectaron individuos de *Pediculus humanus capitis* (piojos), de las cabezas de niños/as entre 6 y 12 años. Los ejemplares fueron separados entre hembras y machos, y posteriormente se realizaron diferentes fases de procesamiento para la extracción de exocutícula, mediante un lavado externo y para la extracción de intestino se realizaron las disecciones por individuo y luego se realizó un homogéneo por intestino. Cada preparado (exocutícula o intestino), se homogeniza, para luego ser sembrado en placas de Petri de 90 mm, y se dejan en estufa a 27-28 °C, por al menos 7 y hasta 15 días. De estos 237 individuos muestreados, solo 98 presentaron crecimientos fúngicos, de los cuales 78 fueron en hembras y 19 en machos. Del total que presentaron crecimiento, se extrajeron 184 cepas totales, 167 provienen de hembras y 17 de machos. Se realizó un análisis de poisson para determinar en qué sitio de extracción hubo mayor crecimiento fúngico, pero se observó una sobredispersión debido al desbalance que presentaba el modelo, por lo que se realizó un análisis de binomial negativa, para intentar solucionar la sobredispersión. Los resultados reflejaron que hay mayor presencia de cepas en piojos hembras sobre piojos machos, mientras que en exocutícula e intestino los crecimientos fúngicos fueron similares dentro de un mismo sexo.

Palabras clave: **niños, piojos, hongos**

DIFERENCIACIÓN DE VARIEDADES DE CANNABIS SATIVA SEGÚN LA COMPOSICIÓN DE CANNABINOIDES EMPLEANDO ANÁLISIS MULTIVARIADO

LEONARDO BAJDA^{1,2}, VERÓNICA LAC PRUGENT², MARÍA MARCELA AMARO¹

¹*Instituto de Investigación y Desarrollo en Ingeniería de Procesos, Biotecnología y Energías Alternativas, CONICET - Universidad Nacional del Comahue*

²*Departamento de Estadística, Facultad de Economía y Administración, Universidad Nacional del Comahue*

lacprugentvero@gmail.com

RESUMEN

Durante los últimos cuatro años, el Laboratorio de Cromatografía del PROBIEN ha proporcionado servicios de análisis de cannabinoides en muestras de aceites medicinales de *Cannabis sativa*. En este período, se ha recopilado una base de datos extensa con información sobre diversas variedades empleadas por los usuarios. El objetivo de esta investigación es identificar, a través de un análisis multivariado, grupos de variedades de cannabis con perfiles de cannabinoides similares. Esto simplificaría la sustitución de una variedad por otra, asegurando efectos terapéuticos equivalentes. Para la identificación y cuantificación de los cannabinoides analizados, se utilizó un cromatógrafo HPLC-DAD y se emplearon los estándares analíticos de tetrahidrocannabinol (THC), cannabidiol (CBD), cannabigerol (CBG) y cannabinol (CBN). Para el estudio estadístico de los datos, se llevó a cabo un análisis de componentes principales (ACP) para reducir la dimensionalidad e identificar gradientes, seguido de un análisis de conglomerados jerárquico utilizando la distancia de Jaccard y el método de encadenamiento completo. En el ACP se observó un gradiente creciente en el eje horizontal para el CBD, un gradiente decreciente en el mismo eje para el THC, un gradiente creciente en el eje vertical para el CBG y un gradiente oblicuo, decreciente en el eje horizontal y creciente en el eje vertical, para el CBN. El análisis de conglomerados se empleó para agrupar las variedades en clústeres basados en sus perfiles de composición. Esta técnica permitió formar grupos bien definidos, con una correlación cofenética de 0.972. Estos clústeres reflejan las similitudes en la composición relativa de los cannabinoides, facilitando la identificación de las variedades más afines entre sí. Las conclusiones de esta investigación representan un valioso aporte que facilita la selección de variedades adecuadas para cada tratamiento terapéutico, contribuyendo con un gran beneficio a cultivadores, pacientes y médicos.

Palabras clave: ***terapia cannabica, componentes principales, análisis de conglomerados, aceites medicinales.***



ESTIMACIÓN NO DESTRUCTIVA DE ÁREA FOLIAR EN CLONES DE CACAO

FERNANDO CASANOVES^{1,4}, JULIO DI RIENZO², JAIRO ROJAS³, JUAN CARLOS SUAREZ⁴

¹CATIE- Centro Agronómico Tropical de Investigación y Enseñanza, Costa Rica

²Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba

³AGROSAVIA, Corporación Colombiana de Investigación Agropecuaria

⁴Universidad de la Amazonia, Colombia

casanoves@catie.ac.cr

RESUMEN

El monitoreo del crecimiento de los cultivos utiliza modelos matemáticos para describir la relación entre el crecimiento de una planta, la producción de materia seca y la expansión del área foliar total. El estudio a través del tiempo de estas variables permite evaluar los efectos de diferentes condiciones ambientales y de manejo sobre el crecimiento del cultivo. La mayoría de los estudios que determinan el crecimiento en condiciones controladas requieren muestreo destructivo, pero esto es inaplicable en experimentos que apuntan a monitorear el crecimiento de la misma hoja a lo largo del tiempo. Se ajustaron modelos de regresión para predecir el área de las hojas de cacao (*Theobroma cacao L.*) utilizando hojas de 22 clones provenientes de la región cacaotera de Colombia. Se recolectaron un total de 15586 hojas de tal manera que se cubriera un amplio rango de tamaños. Se midió el ancho, largo y área foliar de cada hoja. El número total de hojas se dividió aleatoriamente en conjuntos de entrenamiento y validación. El conjunto de entrenamiento se utilizó para el ajuste y selección del modelo, el otro para medir su capacidad de predicción. El área de las hojas se estimó utilizando diferentes modelos de regresión lineal con base en el largo y el ancho de la hoja. Se evaluaron modelos generales y clon específicos usando los criterios AIC y BIC, y la relación entre los valores observados y predichos por el modelo. Las regresiones polinómicas general y clon específicas que involucraron tanto el largo como el ancho de las hojas proporcionaron buenos modelos de predicción. Los modelos específicos para cada clon tuvieron mejor desempeño en los clones con mayor rango de área foliar explorada.

Palabras clave: **Análisis de regresión múltiple, morfología de la hoja, validación de modelo.**



ESTUDIO COMPARATIVO DE TÉCNICAS ANALÍTICAS PARA LA EPIDEMIOLOGÍA VETERINARIA

NICOLAS FLAIBANI^{*1,2}, TOMAS GONZALEZ GARELLO^{*1,2}, VALENTIN ROZENBLIT^{1,2}, ELISA HELMAN^{**1,3},
MARIANA BERNSTEIN³, LAIS PARDINI³, MA. SOLEDAD FERNÁNDEZ^{1,2}

¹*Grupo de Bioestadística Aplicada, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

²*Instituto de Cálculo, Universidad de Buenos Aires-CONICET*

³*Laboratorio de Inmunoparasitología (LAINPA), Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, La Plata*

*Ambos autores contribuyeron equitativamente

***elisahelman@gmail.com*

RESUMEN

La alfabetización estadística es fundamental para promover la óptima interpretación y aplicación de conceptos estadísticos en contextos científicos. En epidemiología veterinaria, usualmente se recurre a técnicas estadísticas básicas, aun cuando los datos requieren enfoques más avanzados. Por ejemplo, la prevalencia de infecciones se suele analizar con la prueba de Chi-cuadrado, aunque no permite cuantificar la magnitud de la asociación entre el patógeno y otras variables de interés. Este estudio se centró en comparar esta prueba versus un modelo lineal generalizado (GLM), en particular una regresión logística (RL). Asimismo, para entender si la asociación entre variables de estudio es lo suficientemente fuerte como para ser clínicamente relevante y/o para guiar políticas de salud pública, resulta fundamental considerar una medida de magnitud de efecto (p.ej.: odds ratio, OR). Se tomó como ejemplo un estudio sobre la seroprevalencia de *Toxoplasma gondii* en cerdos de Argentina, según el sistema de producción, clasificado en intensivo (CI: animales en confinamiento durante su ciclo productivo) y semiextensivo (CSE: animales de producción familiar, criados al aire libre). Se analizaron 593 muestras serológicas independientes (CI=272 y CSE=321). Los resultados mostraron diferencias significativas para el tipo de cría, con ambas técnicas (p -valor < 0.05). La prueba de Chi-cuadrado indicó que la prevalencia no es independiente de dicha variable. En cambio, la RL proporcionó una cuantificación del efecto con un OR de 27 ($IC_{95\%} = 17-42.4$). Esto significa que las chances de detectar *T. gondii* son 26 veces mayores en el CSE en comparación con el CI. Este trabajo destaca la importancia de utilizar las técnicas estadísticas más adecuadas para maximizar la información y mejorar las conclusiones. En resumen, los GLM permiten modelar probabilidades y relaciones complejas, mientras que el Chi-cuadrado solo evalúa asociaciones categóricas sin medir probabilidades ni ajustar múltiples variables.

Palabras clave: **Alfabetización estadística, Salud Pública Veterinaria, Test de independencia, Modelos lineales generalizados.**

APLICACIÓN DE MODELOS DE REGRESIÓN LOGÍSTICA PARA EVALUAR EL IMPACTO DE DIFERENTES FACTORES SOBRE LA TASA DE PREÑEZ EN VACAS

MUNILLA, S¹² y DE LOYOLA, J.M.¹

¹Facultad de Agronomía, Universidad de Buenos Aires

²Instituto de Investigaciones en Producción Animal (INPA - CONICET)

munilla@agro.uba.ar

RESUMEN

La cría bovina es una actividad económica cuyo objetivo es la producción de terneros sobre la base de un capital de explotación vivo: la vaca. Una vez alcanzada la madurez sexual, una vaca puede producir un ternero por año, y este es el objetivo productivo que los ganaderos se fijan. La eficiencia reproductiva de un rodeo se evalúa mediante la tasa de preñez, que representa el porcentaje de vacas preñadas luego de la temporada de servicios (apareamientos). Al estudiar los factores que influyen sobre la tasa de preñez, es común evaluarlos individualmente mediante pruebas de Chi-cuadrado. Sin embargo, este trabajo ilustra que el uso de modelos de regresión logística es una metodología más adecuada. Se analizaron datos de un rodeo Hereford ajustando un modelo de regresión logística a la condición de cada vaca (preñada o vacía) mediante la función 'glm' del programa R. Como variables predictoras se incluyeron el peso vivo (covariable), la edad (dos niveles: 'joven' y 'adulta') y la condición corporal, una medida del estado nutricional de la vaca (dos niveles: 'buena' y 'deficiente'), todas registradas antes del servicio. Los resultados se alinearon perfectamente con principios zootécnicos bien establecidos: la condición corporal y la edad de la vaca tuvieron efectos significativos ($p < 0,05$) sobre su probabilidad de preñarse. Pero, además, la metodología permitió cuantificar estos efectos. Así, se estimó que las vacas con buena condición corporal antes del servicio tuvieron 2,7 veces más chances de preñarse que aquellas con condición corporal deficiente, y que las vacas adultas tuvieron 2,65 veces más chances de preñarse que las jóvenes. Se concluye que el modelo de regresión logística es una herramienta poderosa para analizar cuantitativamente el efecto de múltiples factores sobre la tasa de preñez en vacas de cría.

Palabras clave: **fertilidad bovina, variables binarias, análisis multifactorial.**

MODELOS DE REGRESIÓN LOGÍSTICA PARA PREDECIR LA HABILIDAD DE PERMANENCIA EN HEMBRAS BOVINAS CRIOLLAS

TOPAYAN MV^{1,2,3}, MARTÍNEZ RD^{1,2}, MUNILLA S^{4,5}

¹Facultad de Ciencias Agrarias, Universidad Nacional de Lomas de Zamora

²Instituto de Investigación sobre Producción Agropecuaria, Ambiente y Salud (IIPAAS CIC)

³Comisión Nacional de Investigaciones Científicas y Técnicas (CONICET)

⁴Facultad de Agronomía, Universidad de Buenos Aires

⁵Instituto de Investigaciones en Producción Animal (INPA - CONICET)

vtopayan@agrarias.unlz.edu.ar

RESUMEN

La habilidad de permanencia es una medida clave de la eficiencia reproductiva en hembras bovinas y se define como la probabilidad de que una vaca permanezca en el rodeo hasta alcanzar un retorno económico favorable, comúnmente asociado con seis partos consecutivos. Este atributo se registra de manera binaria: éxito o fracaso, lo que hace que los modelos de regresión logística (MRL) sean adecuados para su análisis. Se sabe que las condiciones ambientales al momento del nacimiento y del primer servicio de una ternera, especialmente las precipitaciones, que influyen en la disponibilidad de forraje, pueden afectar su habilidad de permanencia futura. Asimismo, la edad en la que una vaquillona recibe su primer servicio también puede ser importante, particularmente en servicios tempranos. Este estudio tuvo como objetivo evaluar cuál o cuáles de estos factores pueden utilizarse para predecir la habilidad de permanencia de hembras bovinas. Para ello se analizaron mediante MRL datos de un rodeo de vacas Criollas nacidas entre 1989 y 2021, sometidas a un servicio a los 15 meses. Se utilizaron registros de precipitaciones para definir las condiciones climáticas del año de nacimiento y del primer servicio, clasificándolos en seco (<827 mm), normal (827–1323 mm) o húmedo (>1323 mm). Además, se incorporó la edad al primer parto (EPP) como covariable. La comparación entre los modelos se realizó mediante el criterio de “*leave-one-out cross-validation*”. El modelo con mayor exactitud fue el que sólo incluyó a la EPP (exactitud = 0,615). Por cada mes adicional en la EPP, en el rango que caracteriza a un servicio temprano, las chances de lograr seis partos consecutivos se duplicaron (OR = 2,074, IC_{95%} = [1,178 – 3,796], *valor-p* = 0,014).

Palabras clave: **eficiencia reproductiva, bovinos para carne, servicio temprano**



INTERACCIÓN ENTRE FACTORES AMBIENTALES PARA ENTENDER LA RESPUESTA DE LA RIQUEZA DE AVES A GRADIENTES URBANO-RURALES

ANAHÍ SOFÍA VACCARO¹; JULIETA FILLOY¹

¹*Departamento de Ecología, Genética y Evolución, IEGEBA (CONICET-UBA), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

anahivaccaro@ege.fcen.uba.ar

RESUMEN

La riqueza de aves suele aumentar con la productividad primaria, que presenta estacionalidad en climas templados. Esta productividad varía a lo largo del gradiente urbano-rural, pero el patrón y su intensidad dependen del contexto de uso del suelo y del bioma. Este estudio compara la riqueza de aves entre sectores del gradiente en dos contextos de uso de la tierra: agricultura con cultivos anuales y ganadería en pastizales seminaturales, en primavera e invierno. Se espera que la riqueza de aves sea mínima en los centros urbanos, intermedia en la periferia y máxima en el área rural, siendo mayor en primavera en el contexto ganadero y en invierno en el agrícola. Se seleccionaron 10 ciudades de tamaño similar en la región Pampeana, cinco en cada contexto. Se relevaron las aves en cinco puntos en cada sector del gradiente: centro urbano, periferia y rural, registrando la riqueza de aves por sector y ciudad en cada estación. Se estudiaron modelos lineales generalizados mixtos con distribución Poisson (función glmer del paquete lme4), considerando la riqueza como variable respuesta y el sector, contexto y estación como factores fijos, incluyendo interacciones dobles y triple, con ciudad (anidada en la región) como factor aleatorio. El mejor modelo, según el criterio de información de Akaike, incluyó solo la interacción entre estación y contexto. Se cumplió el supuesto de homogeneidad de varianzas, aunque no de normalidad, pero se confirmó la ausencia de sub y sobredispersión. La riqueza fue un 158% mayor en el área rural y un 58% mayor en la periferia en comparación con el centro urbano. En contexto ganadero respecto al agrícola, la riqueza fue un 32% mayor en primavera, mientras que en invierno fue un 17% menor. El uso agrícola de la tierra atenúa la estacionalidad de la productividad, influyendo en la riqueza de aves.

Palabras clave: *estacionalidad, glmm, riqueza de aves, uso de la tierra.*

DISFUNCIONES DEL PISO PÉLVICO EN MUJERES ARGENTINAS QUE REALIZAN ACTIVIDAD FÍSICA Y SU IMPACTO FUNCIONAL. ESTUDIO EPIDEMIOLÓGICO, ANALITICO

TEDESCO FRANCO¹, RUGGIERO NOELIA², LIBRYK AGUSTIN²

¹Sanatorio Güemes, Ciudad Autónoma de Buenos Aires

²Hospital Zonal General de Agudos Petrona V. de Cordero, Provincia de Buenos Aires

fran07tedesco@gmail.com

RESUMEN

La incontinencia urinaria (IU) y anal (IA) en mujeres deportistas restringen la participación en la actividad física. El reconocimiento y abordaje adecuado de los factores relacionados con estas deficiencias es necesario. Nuestro objetivo es reportar la frecuencia de aparición de disfunciones de piso pélvico e identificar los factores asociadas a estas disfunciones. Se llevó a cabo una encuesta auto reportada. Se incluyó un total de 287 mujeres mayores de edad que practicaban actividad física regularmente. Se describen las características generales y por grupos de incontinencia. Se realizaron comparaciones univariadas para las variables explicativas de interés. Dada la escasa bibliografía sobre el efecto del nivel de entrenamiento y características demográficas en nuestra región, se realizó un análisis exploratorio de las variables explicativas que diferencian a los sujetos con incontinencia utilizando el método Random Forest. Además, se construyó un modelo de Regresión Logística multivariada para determinar el Odds Ratio de cada variable. El 68% presentaron síntomas de IU y el 33% de IA. En el análisis univariado, los sujetos con IU tenían un Índice de masa corporal (IMC) significativamente mayor, y para IA, una mayor edad. Con el método Random forest, la tasa de error fue del 33.1% y 34.5%. Las variables de mayor relevancia para IU y UA fueron IMC, años de práctica deportiva previa y Edad. En cuanto al modelo de regresión logística multivariada para IU, la única variable con significancia estadística del test de Wald fue IMC OR=1.11(1.02;1.21). Para IA, lo fueron Cirugía pélvica OR=3.10 (1.25;7.77 p-v=0.01), Menopausia OR=4.61 (1.22;17.88, p-v=0.02), Edad OR=0.94 (0.90;0.98 p-v=0.01) y Años de deportes previos OR=1.04 (1.007;1.07 p-v=0.01). La variable años de práctica previa no fue significativa en ninguno de los casos de IA OR 0.99 (0.96;1.03 p-v=0.86) e IU OR=0.99 (0.96;1.03, p-v=0.85). El actual es el primer estudio de disfunciones del piso pélvico y sus relaciones con la actividad física en Argentina.

Palabras clave: ***Piso pélvico, Actividad física, Incontinencia, Random forest, Regresión logística.***

RESTAURACIÓN DE PASTIZALES INVADIDOS: EVALUACIÓN EXPERIMENTAL DE TÉCNICAS DE MANEJO CON DISEÑO FACTORIAL

CAGNONE MF^{1,2}, CRISTIANO PM^{1,2}, DIAZ VILLA MVE^{1,2}, TOGNETTI PM³

¹Laboratorio de Ecología Funcional, Departamento de Ecología Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires.

² Instituto de Ecología, Genética y Evolución (IEGEB), Universidad de Buenos Aires-CONICET

³ Instituto de Investigaciones Fisiológicas y Ecológicas Vinculadas a la Agricultura, Facultad de Agronomía, CONICET-Universidad de Buenos Aires (IFEVA-CONICET-UBA).

mcagnone@ege.fcen.uba.ar

RESUMEN

El control de plantas invasoras es un paso crucial en la restauración ecológica. El diseño de experimentos a campo permite evaluar estrategias de manejo de manera efectiva. En este estudio, se evaluaron estrategias para el control de plantas leñosas invasoras y la recuperación de especies nativas en un pastizal pampeano (Campana, Buenos Aires). Se realizó un experimento con diseño factorial en parcelas divididas y replicado en seis bloques (primavera 2023). Cada bloque consistió en cuatro parcelas de 8x8 m², donde se combinó un tratamiento de disturbio del suelo (sin/con arado) con la aplicación de herbicida de hoja ancha (sin/con 2,4-D). En cada parcela se establecieron dos subparcelas circulares de 3 m² sin o con siembra de gramíneas nativas. La abundancia-cobertura de suelo desnudo (SD), gramíneas nativas (GN) y leñosas invasoras (LI) se estimó en las subparcelas en dos fechas (diciembre 2023 y marzo 2024), para mejorar la identificación de las especies. Los valores de cada variable respuesta fueron analizados mediante modelos lineales generalizados mixtos con distribución beta (enlace logit). En cada modelo, los factores predictores fijos fueron disturbio, herbicida y siembra, con sus interacciones. Los factores aleatorios fueron la parcela anidada dentro del bloque, y la fecha. La significancia de cada término se evaluó mediante tests de Wald tipo II ($\alpha=0,05$). No hubo interacciones significativas para SD y GN. El SD aumentó 5,9% con el disturbio y disminuyó 5,0% con la siembra. Las GN aumentaron 9,2% con el herbicida. Para las LI, las interacciones herbicida*disturbio y herbicida*siembra fueron significativas. Las LI aumentaron con el herbicida 4,6% cuando no hubo disturbio y 3,8% cuando hubo siembra. Sin herbicida, disminuyeron 3,6% con la siembra. La magnitud del efecto de los tratamientos fue baja a corto plazo. Éstos serán re-aplicados pasado un año y se re-evaluarán sus efectos.

Palabras clave: ***leñosas invasoras, disturbio del suelo, herbicida de hoja ancha, siembra de gramíneas nativas, GLMM, distribución beta***

Ciudad de Buenos Aires, octubre 2024

CONSIDERACIÓN DEL DISEÑO DE MUESTREO EN EL ANÁLISIS ESTADÍSTICO DE PATRONES ALTITUDINALES DE RIQUEZA

BOVINA, SOL A.¹; FILLOY, JULIETA¹; RAMOS, CAROLINA S.¹

¹Dpto. de Ecología Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

²Instituto de Ecología, Genética y Evolución (IEGEB), Universidad de Buenos Aires-CONICET

solbovina@gmail.com

RESUMEN

Entender los patrones de distribución de la biodiversidad es un desafío para la ecología, y realizar un análisis adecuado es una tarea para la bioestadística. Se estudiaron los patrones altitudinales de riqueza taxonómica de plantas, en dos montañas en Argentina. Para ello, se ubicaron 10 puntos en bandas cada 100 metros de altitud y en cada sitio se ubicaron dos cuadrantes. Allí se relevaron especies de plantas y variables ambientales, conformando un muestreo con diseño de factores con niveles anidados. El objetivo fue analizar la forma de los patrones altitudinales y estudiar el grado de apoyo a las hipótesis de la energía y de la dinámica agua-energía. En este contexto nos preguntamos si los patrones detectados cambian según como se considere el diseño de muestreo en el análisis de los datos. Se consideraron dos aproximaciones a la variable respuesta: riqueza acumulada por banda y riqueza por cuadrante incorporando los efectos aleatorios banda y punto, se utilizó en ambas una distribución Poisson y se especificaron GLM y GLMM respectivamente. Para estudiar el patrón altitudinal se utilizaron ecuaciones que representan los patrones teóricos y se modeló la riqueza de especies en función de la altura. Para estudiar el apoyo a las hipótesis de energía se modeló la riqueza en función de distintas combinaciones de variables ambientales explicativas. Para ambos análisis la selección de modelos se realizó con el criterio de Akaike. Además, para la aproximación anidada se encontró significancia del factor aleatorio, indicando que su incorporación explica parte de la variabilidad. Los resultados para ambas aproximaciones fueron diferentes respecto a la forma teórica que mejor ajustó al patrón altitudinal y cuáles fueron los mejores modelos en la asociación con las variables ambientales. La aproximación elegida para realizar el análisis estadístico deriva en conclusiones diferentes y en un problema al comparar entre varias investigaciones.

Palabras clave: **Patrones altitudinales – Riqueza – Diseño – Selección de modelos**

INCIDENCIA DEL COEFICIENTE DE FRUTO OCULTO EN LAS ESTIMACIONES DE PRODUCCIÓN DE CÍTRICOS DE LA PROVINCIA DE CORRIENTES

PONCE DE LEÓN, L.¹, CHABBAL, M.¹, HIDALGO, M.¹, GIMÉNEZ, L.¹

¹ *Facultad de Ciencias Agrarias, Universidad Nacional del Nordeste*

luciamatildeponcedeleon@gmail.com

RESUMEN

El número de frutos por árbol (carga) es un componente esencial en el modelo de pronóstico utilizado para estimar volúmenes de producción de cítricos. Cuando la estimación de la carga se lleva a cabo por métodos visuales se puede utilizar un factor de corrección denominado coeficiente de fruto oculto (CFO). El presente trabajo tuvo como objetivo comparar las estimaciones de carga por árbol, con y sin la corrección por fruto oculto. Las estimaciones de carga y la cosecha real se realizaron en huertos comerciales de mandarino tangor Murcott (*Citrus reticulata* x *C. sinensis* 'Murcott') y de naranja dulce (*Citrus sinensis* L.) de las provincias de Corrientes y Entre Ríos entre los años 2007 y 2016. Se trabajó con datos de árboles con más de 10 años. Para determinar el factor de corrección por CFO, cada grupo de datos se dividió al azar en una proporción 70:30, el 70% se utilizó para la obtención de los CFOs y el 30 % restante para la aplicación de dichos coeficientes. A partir del 70% de los datos se obtuvieron los CFO promedio, los cuales fueron utilizados posteriormente en el 30% restante de los datos para corregir las estimaciones. La comparación de las estimaciones de carga (corregida y sin corregir) se realizó a partir de Modelos Lineales Mixtos considerando el huerto comercial como factor aleatorio. La precisión de las estimaciones se evaluó mediante el error porcentual. Tanto en mandarino como en naranjo se observaron diferencias significativas entre las estimaciones con y sin la corrección por CFO, en mandarino el P-valor=0.0297 y en naranjo P-valor<0.0001. Las estimaciones de carga corregidas tuvieron levemente mayor error porcentual que las estimaciones sin corregir en mandarino, mientras que en naranjo las estimaciones corregidas tuvieron menos error que las estimaciones sin corregir.

Palabras clave: ***Cítricos, Estimación de carga, Coeficiente de fruto oculto.***

IMPACTO EN LA ESTIMACIÓN DE LA MAGNITUD DEL EFECTO Y EN LA PROBABILIDAD DE ERROR DE TIPO I, AL OMITIR LA ESTRUCTURA JERÁRQUICA EN UN MODELO LINEAL GENERALIZADO MIXTO

ELISA HELMAN^{1,3}, ADRIANA PEREZ¹, PABLO TURJANSKI^{1,4}, JUAN MANUEL UNZAGA³, MA. SOLEDAD FERNÁNDEZ^{1,2}

¹Grupo de Bioestadística Aplicada, Fac.de Ciencias Exactas y Naturales, Universidad de Buenos Aires

²Instituto de Cálculo, (Universidad de Buenos Aires-CONICET)

³Laboratorio de Inmunoparasitología (LAINPA), Facultad de Ciencias Veterinarias, Universidad Nacional de La Plata, La Plata

⁴Instituto de Ciencias de la Computación, Universidad de Buenos Aires-CONICET

elisahelman@gmail.com

RESUMEN

En epidemiología veterinaria los abordajes analíticos suelen ignorar la estructura jerárquica de los datos, como por ejemplo la de animales criados en establecimientos (granjas, corrales, tambos, etc.) Este trabajo evalúa cómo omitir la estructura jerárquica de los datos afecta la magnitud del efecto y la probabilidad de error tipo I. Se simuló datos de presencia/ausencia de *Sarcocystis* spp. en cerdos agrupados en granjas (100 establecimientos, 50 cerdos por establecimiento), usando un modelo lineal generalizado mixto (GLMM), con dos variables predictoras fijas: una a nivel de granja (sistema de cría: intensivo y semi-extensivo) y otra a nivel de individuo (edad: lechón y cerdo destetado), además de granja como efecto aleatorio. Se plantearon distintos escenarios de valores de varianza entre granjas (σ^2), de efecto del sistema de cría (β_1) y de la edad (β_2). En cada proceso se realizaron 250 simulaciones bajo diseños balanceados. Los resultados de estas simulaciones se compararon con otras en las que cerdos de una misma granja se consideraron independientes (GLM). Se encontró que al ajustar modelos GLM, cuanto mayor es la varianza entre granjas, mayor es la subestimación de las estimaciones de β_1 y β_2 (hasta un 45%, en el rango estudiado). Además, esta subestimación aumenta a medida que se incrementa la magnitud del efecto. Respecto a las pruebas para detectar significación de β_1 , se encontró un aumento de la probabilidad de cometer error de tipo I de hasta 75% para β_1 en escenarios de alta varianza del factor aleatorio, mientras que para β_2 no se encontró un incremento. Este trabajo destaca la importancia de incorporar el diseño experimental o muestral en el modelado de datos, de manera de evitar pseudoreplicación y aumento de la probabilidad de error tipo I, así como sesgos en las estimaciones de los parámetros.

Palabras clave: **Datos jerárquicos - Pseudoreplicación - Modelos lineales generalizados mixtos - Error de tipo I - Magnitud de efecto**



Modelos No Lineales

APLICACIÓN Y ANÁLISIS DE LAS CURVAS DE PERFORMANCE TÉRMICA EN EL ESTUDIO DEL NICHO ECOLÓGICO DE ESPECIES DE DROSOPHILA

LUCAS KREIMAN¹, PABLO SCHILMAN², JULIÁN MENSCH¹

¹Laboratorio de Evolución, Departamento de Ecología, Genética y Evolución, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

²Laboratorio de Ecofisiología de Insectos, Departamento de Biodiversidad y Biología Experimental, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

kreimanlucas@gmail.com

RESUMEN

Las curvas de performance térmica son una herramienta teórica de amplio uso en la biología que describe matemáticamente la relación con la temperatura de diferentes caracteres biológicos al representar la *performance* en un eje vertical contra la temperatura, variable independiente, en el eje horizontal. Son utilizadas para realizar comparaciones interespecíficas del nicho térmico, así como para asestar el efecto de la dieta y otros componentes ambientales sobre la performance de una especie a lo ancho de todo su rango de tolerancia térmica, en oposición a una única temperatura de referencia (i. e. 25°C). La performance puede estar representada por variables como las tasas de desarrollo, crecimiento y asimilación, la fecundidad y la supervivencia, o incluso caracteres biológicos de los más variados, como actividad sensorial o inmunológica. La diversidad de variables abarca varios niveles de organización y una amplia diversidad de distribuciones de probabilidad estadística. En los últimos años se ha multiplicado el número de modelos matemáticos utilizados para ajustar a estas curvas, haciendo de la selección de modelos un desafío en sí mismo. En este trabajo abordaremos un *pipeline* ampliamente difundido en la literatura para la selección de modelos y la comparación estadística de las curvas de performance. Utilizaremos como caso de estudio el *fitness* de *Drosophila buzzatii* (Diptera: Drosophilidae), estimado a partir de caracteres morfológicos y de historia de vida, sobre su rango térmico en hospedadores y dieta de laboratorio. Respecto a otras especies, *D. buzzatii* presenta una curva euritérmica, ancha y baja, correspondiente a una estrategia térmica generalista. Si bien las curvas no difieren en su ancho térmico ni en su temperatura óptima, las performance máximas difieren según la dieta de cría utilizada respondiendo como era esperado según la jerarquía de elección de hospedadores en la naturaleza.

Palabras clave: *rTPC*, *Drosophila buzzatii*, *biología térmica*

Ciudad de Buenos Aires, octubre 2024

SELECCIÓN DE MODELOS DE CRECIMIENTO POBLACIONAL EN *Daphnia pulicaria*: UNA EVALUACIÓN MEDIANTE NLS Y AIC

DANIEL NINO FLORES-MENDEZ^{1,2}, MARÍA FLORENCIA GUTIERREZ^{2,3}, MELINA CAMPERO¹

¹Unidad de Limnología y Recursos Acuáticos, Universidad Mayor de San Simón, Cochabamba, Bolivia

²Instituto Nacional de Limnología (CONICET-UNL), Universidad Nacional del Litoral, Santa Fe

³Escuela Superior de Sanidad "Dr. Ramón Carrillo", Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Santa Fe

d.floresm@santafe-conicet.gov.ar

RESUMEN

La aplicación de herramientas estadísticas avanzadas es importante para la correcta interpretación de datos ecológicos. A nivel poblacional, el crecimiento se estudia generalmente mediante la estimación de tasas demográficas, como la tasa de crecimiento poblacional (r) y, en algunos casos, su capacidad de carga (K), las cuales pueden estar condicionadas por factores ambientales. Sin embargo, existe una variedad de modelos matemáticos para el cálculo de estos parámetros, y una selección inadecuada puede llevar a interpretaciones erróneas. Este trabajo analiza la aplicación de modelos de Mínimos Cuadrados No Lineales (nls) y la selección de modelos mediante el Criterio de Información de Akaike (AIC) para elegir los modelos de crecimiento poblacional en *Daphnia pulicaria*, una especie de microinvertebrado acuático, sometida a diferentes condiciones de temperatura (baja, media, alta) y radiación lumínica (PAR, PAR+UV). Evaluamos siete modelos de crecimiento poblacional: denso-independiente (M1), denso-dependientes (M2, M3, M4, M5), y denso-dependientes con retraso temporal (M6, M7). Los parámetros de crecimiento mostraron un rango de variación de hasta 16 veces para r y el doble para K entre los 7 modelos. Los mejores modelos de crecimiento según el AIC dependieron de las condiciones de temperatura y luz: M6 fue el mejor para poblaciones a temperaturas bajas y medias con luz PAR y a temperaturas altas con PAR+UV, mientras que M7 fue el más adecuado a temperaturas bajas y medias con luz PAR+UV y temperaturas altas con PAR. Estos resultados subrayan la importancia de seleccionar el modelo adecuado para interpretar correctamente la dinámica poblacional bajo diferentes condiciones ambientales. La elección precisa del modelo no solo mejora la comprensión del impacto de la temperatura y la luz en el crecimiento de estos organismos, sino que también optimiza las predicciones de cambios en estas poblaciones frente a las fluctuaciones ambientales.

Palabras clave: **Modelos no anidados, Parámetros de crecimiento, Cladóceros, Poblaciones de alta montaña**

SELECCIÓN DE PARÁMETROS PARA DESCRIBIR EL CRECIMIENTO EN FRUTOS DE MANDARINO TANGOR MEDIANTE MODELOS NO LINEALES MIXTOS

RODRÍGUEZ AZCONA ROCIO INES¹, MENDOZA MARIA ELIZABETH¹, CARDOZO ORNELLA ABRIL¹,
BÓBEDA GRISELDA RITA ROMINA¹,

¹ Cátedra Cálculo Estadístico y Biometría, Departamento Matemática, Facultad de Ciencias Agrarias, Universidad Nacional del Nordeste, Corrientes.

rociorodriguezazcona@gmail.com

RESUMEN

Los pronósticos de cosecha de fruta son de gran utilidad para planificar la cosecha. Para ello es importante contar con información acerca del crecimiento de los frutos a lo largo del tiempo. El objetivo de este trabajo fue evaluar modelos no lineales mixtos y seleccionar los parámetros que tienen mejor capacidad predictiva para describir el crecimiento de frutos de tangor Murcott en las provincias de Corrientes y Entre Ríos, y seleccionar el modelo con mayor capacidad predictiva. Se recopiló información de huertos comerciales de mandarina, y se seleccionaron 10 árboles mediante un muestreo sistemático con arranque aleatorio, resultando en total 182 combinaciones (TemporadaxHuertoxArbol). En cada árbol se identificó 30 frutos y se le midió el diámetro ecuatorial (DE) en mm, con calibre digital, con una frecuencia de 10-15 días. Se evaluaron modelos no lineales mixtos para describir las curvas de crecimiento sigmoideas: el modelo Gompertz y la quinta reparametrización del modelo Logístico (L5), considerando efectos fijos y aleatorios asociados a cada parámetro. Los modelos se ajustaron mediante el método de REML, utilizando la librería nlme del software RStudio. Se determinó que el modelo Gompertz con parámetros alfa y beta aleatorios estima la mayor cantidad de combinaciones (172 de 182). La precisión de cada terna de parámetros se evaluó utilizando la función nls y se realizó una validación cruzada, empleando el 75% de los datos para la estimación del modelo y el 25% restante para evaluar su capacidad predictiva y validar los mismos. A través de la comparación de los criterios MAE, RMSE y CME, se determinó que la terna de parámetros estimados que presentó mayor precisión, con el menor porcentaje de error de estimación, fue: alfa = 76.83194, beta = 1.41962 y gamma = 0.01419, resultando ser la adecuada para estimar el diámetro ecuatorial en función de los días después de plena flor.

Palabras clave: ***modelos no lineales mixtos, validación cruzada, criterios de selección, efectos aleatorios.***



Trabajos Extendidos



ANÁLISIS ESTADÍSTICO DE LA DINÁMICA DE LA INFECCIÓN POR VIRUS SINCICIAL RESPIRATORIO Y ESTIMACIÓN DE SU INCIDENCIA EN EL AMBA

YAMILA ALEN¹, DANIELA PARADA^{1,2}, JULIA DVORKIN^{2,3}

¹CONICET - Universidad de Buenos Aires. Instituto de Cálculo. Buenos Aires

²CONICET

³Centro INFANT de Medicina Traslacional (CEMIT), Universidad Nacional de San Martín

yami.alen@gmail.com

RESUMEN

A nivel mundial, las infecciones respiratorias agudas bajas (IRAB) son la principal causa de admisiones y hospitalizaciones en unidades de cuidados intensivos pediátricas. En particular, el virus sincicial respiratorio (VSR) es el principal agente etiológico en niños menores de 5 años, incidiendo significativamente en los servicios de salud. El objetivo de este trabajo fue realizar un análisis estadístico descriptivo y estimativo sobre la incidencia de las IRAB por VSR en internaciones pediátricas en el Área Metropolitana de Buenos Aires (AMBA), considerando en particular el impacto de la pandemia COVID-19 en la dinámica de las hospitalizaciones. Con datos proporcionados por el CIMeT, se analizaron 12.767 casos de IRAB pediátricas registrados entre 2018 y 2023 en cuatro hospitales públicos del AMBA. La seroprevalencia de VSR se estimó mediante un diseño observacional, prospectivo y multicéntrico, cuyos detalles metodológicos no se desarrollan en esta instancia por razones de extensión, mientras que para la incidencia se empleó un estudio de cohorte retrospectivo, multicéntrico, en una región pediátrica bien definida. Los resultados cuantifican la carga anual de IRAB severa y crítica por VSR en los periodos estudiados. También se evidencia un retraso del pico de casos en 2021-2022, que puede deberse a las medidas de confinamiento y restricciones de circulación por COVID-19, pero en 2023 retorna a una estacionalidad más temprana, posiblemente por cambios en la dinámica de la infección. A partir de datos poblacionales, se estimó la incidencia según grupo etario y año calendario. Con el aumento de la edad los casos disminuyen independientemente del período analizado. Por otro lado, el análisis mostró el efecto de la pandemia, observándose cambios pre y post pandemia en las tasas de incidencia (IRAB y VSR) según rango etario. Los resultados de este análisis son de gran interés para el desarrollo de políticas públicas de salud, siendo un estudio pionero sobre la incidencia de VSR a nivel poblacional en el AMBA.

Palabras clave: ***virus sincicial respiratorio, estadística descriptiva, estimación, incidencia, seroprevalencia, COVID-19***

Ciudad de Buenos Aires, octubre 2024

INTRODUCCIÓN

Las infecciones respiratorias agudas bajas (IRAB) son una de las principales causas de hospitalización e ingreso a unidades de cuidados intensivos pediátricos (UCI) a nivel mundial en niños menores de 5 años (DNE, 2019; Li et al., 2020). Su diagnóstico es clínico y en la mayoría de los casos, se trata de enfermedades leves donde la presencia de factores de riesgo biológicos y socioambientales implican una mayor frecuencia de complicaciones y muerte (Moreno, Barría & Ferrero, 2021). En Argentina, cada año se observa un aumento de los casos de infecciones respiratorias agudas y de la demanda de atención médica y hospitalizaciones (DNE, 2020). El Virus Sincicial Respiratorio (VSR) es el principal agente causal de las IRAB en niños menores de 5 años (DNE, 2020). Esta enfermedad es una importante causa de mortalidad infantil, particularmente en el mundo en desarrollo (Nakajo & Nishiura, 2023) impactando de manera sustancial en los servicios de salud (Li et al., 2020). En países de climas templados como Argentina, la actividad de VSR es mayor durante los períodos de humedad moderada con brotes estacionales durante el otoño e invierno (Lucion & Gentile, 2021). Dada la elevada prevalencia de las IRAB, el análisis estadístico de los casos se presenta como una herramienta valiosa para enfrentarlas y asegurar un uso racional de recursos o para reforzar políticas sanitarias de prevención. Un desafío particular consiste en evaluar los cambios debidos a la pandemia de COVID-19, considerando el impacto de las restricciones sanitarias en la evolución de las infecciones por VSR (Foley et al, 2021). En este sentido, el objetivo principal de este estudio fue analizar la dinámica poblacional de la infección por VSR severa y moderada en menores de 5 años en el Área Metropolitana de Buenos Aires (AMBA, Ciudad Autónoma de Buenos Aires y 40 partidos adyacentes de la Provincia de Buenos Aires). Para ello se exploró la evolución poblacional de la infección por VSR leve y severa en menores de 5 años en el AMBA mediante un análisis descriptivo, y se evaluó el impacto de la pandemia de COVID-19 en la carga anual de IRAB severa y crítica debido al VSR tras la liberación de las restricciones sanitarias públicas desde mediados de 2021 hasta 2023, comparándola con años anteriores (2018-2019).

METODOLOGÍA

Para el análisis, la base de datos fue de 12.767 casos de IRAB pediátricas en cuatro hospitales públicos (Red Infant-ReMBA): Hospital de Niños Pedro Elizalde (Ciudad de Buenos Aires), Hospital Nacional Dr. Alejandro Posadas (El Palomar, oeste del AMBA), Hospital El Cruce (Florencio Varela, sur del AMBA) y Hospital de Niños Sor María Ludovica (La Plata, extremo sur del AMBA). Se tomó como criterio de inclusión a todos los niños menores de 60 meses que ingresaron en uno de estos hospitales entre el 1 de enero de 2018 y el 31 de diciembre de 2023 con diagnóstico clínico de IRAB, excluyendo aquellos casos que fueron dados de alta dentro de las 24 horas sin toma de muestra para diagnóstico viral. El análisis estadístico se realizó empleando R Studio versión 4.4.1. Las variables registradas incluyeron edad, hospitalización, ingreso a UCI, diagnóstico confirmado de VSR, entre otras. Para la estimación de la tasa de incidencia (relación entre casos registrados y población infantil según edades) se recurrió a datos de los Censos Poblacionales 2010 y 2020 del Instituto Nacional de Estadística y Censos (INDEC), y de las proyecciones de los años 2018-2023 basadas en el censo 2010 del INDEC.

RESULTADOS Y DISCUSIÓN

Ciudad de Buenos Aires, octubre 2024

En la Figura 1 se puede observar la cantidad de casos de niños infectados con VSR en función de la semana del año. Se evidencia un desplazamiento de los casos en las semanas de los años 2021 y 2022, donde los picos (moda anual) comenzaron más tarde de lo habitual, probablemente debido a las medidas de restricción y control de la pandemia de COVID-19. Sin embargo, en el año 2023 se observa un adelanto del pico de casos, retornando a una estacionalidad más temprana, lo que sugiere un cambio en la dinámica de la infección post-pandemia.

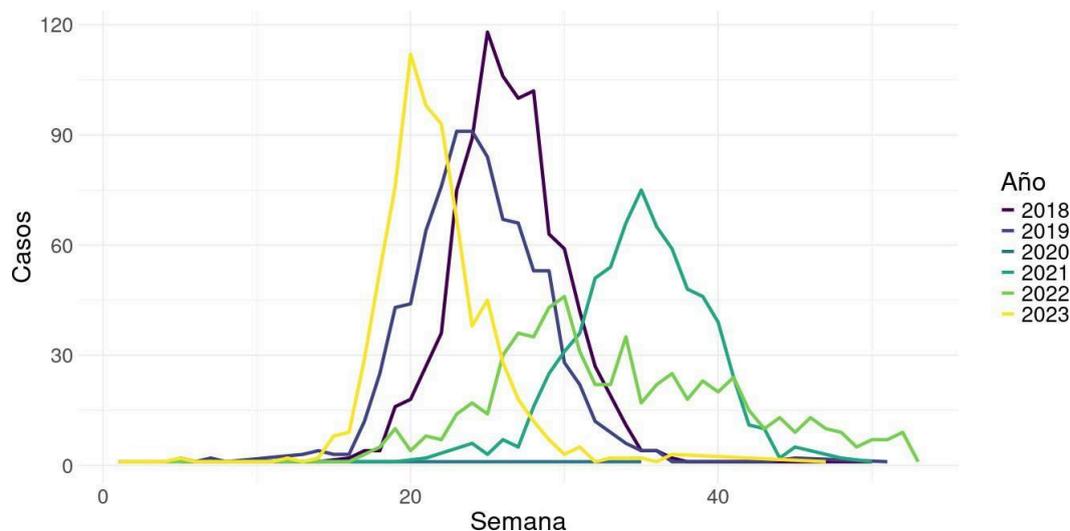


Figura 1: Número de casos en función de la semana (1-52) del año calendario.

En la Figura 2 se presentan los casos por grupo etario y por año, diferenciando entre aquellos relacionados con VSR y los que corresponden a otras IRAB. Esta representación permite apreciar cómo el número de casos varía según la edad, destacando la carga significativa de las IRAB y en particular del VSR en los grupos de menor edad (0-12 y 12-24 meses), independientemente del año. También se observa un marcado descenso de los casos tanto de IRAB como de VSR, sobre todo en la franja etaria de menores a 24 meses, en los años 2020 y 2021 (pandemia COVID19). En 2022 y 2023 se presenta un incremento en las hospitalizaciones, siendo este número siempre menor que en los años pre pandemia.

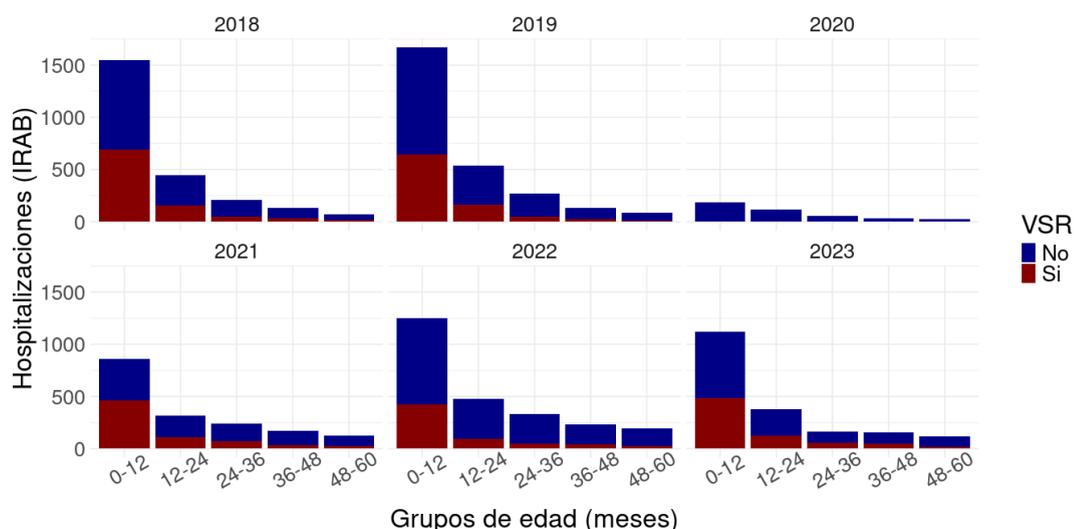


Figura 2: Cantidad de casos de IRAB con VSR (bordó) y sin VSR (azul) según grupo etario por año

En la Figura 3 se muestra la estimación de la tasa de incidencia del VSR en función de grupo etario de acuerdo al año. Se puede apreciar que los grupos más jóvenes (0-12 meses) presentan una mayor incidencia, lo que resalta la vulnerabilidad de esta población. Al aumentar la edad la incidencia disminuye independientemente del año evaluado.

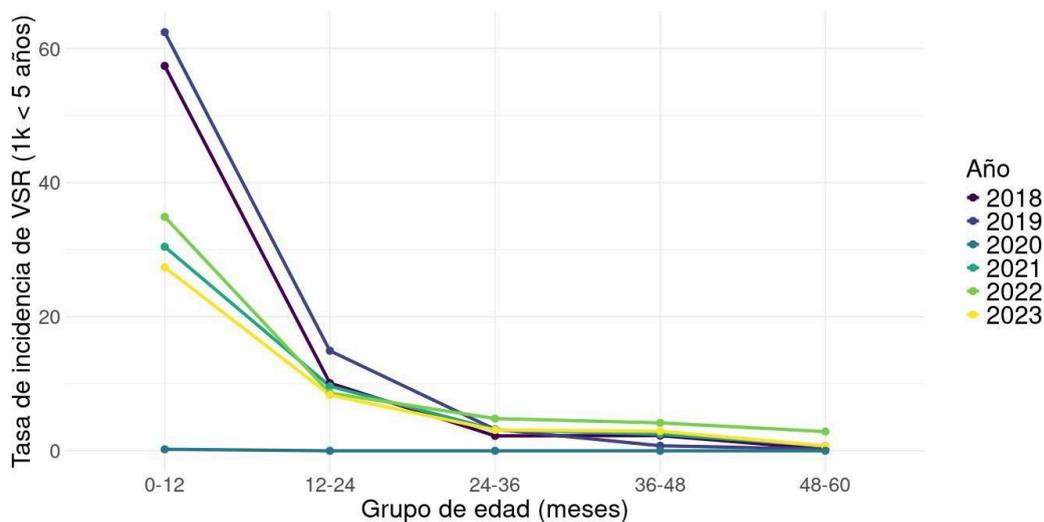


Figura 3: Estimación de la tasa de incidencia de VSR cada 1000 habitantes. por rango etario y año

Por otro lado, el análisis de la Figura 3 permite visualizar el efecto de la pandemia, observándose cambios pre-post COVID en la tasa de incidencia según rango etario. Si comparamos los gráficos de los años pre-pandemia (2018-2019) con los post-pandemia (2021-2023) se puede ver que para las edades más bajas la tasa de incidencia es mayor en los años pre-pandemia que en los de pandemia y post pandemia, mientras que en los grupos de mayor edad ocurre a la inversa.

El análisis realizado mostró el efecto de la pandemia, posiblemente debido a las medidas de confinamiento y restricción implementadas durante la misma, observándose cambios pre y post pandemia en el número de casos y en las tasas de incidencia (IRAB y VSR) según rango etario. Los resultados de este análisis son de gran interés para el desarrollo de políticas públicas de salud, siendo un estudio pionero sobre la incidencia de VSR a nivel poblacional en el AMBA.

REFERENCIAS

- Moreno, L., Barría, S. & Ferrero, F. (2021). *Recomendaciones para el manejo de las infecciones respiratorias agudas bajas en menores de 2 años. Consideraciones para el primer nivel de atención*. Archivos Argentinos de Pediatría, 2021;119(5):292-293.
- Dirección Nacional de Epidemiología y Análisis de la Situación de Salud (DNE) (2019). *Boletín Integrado de Vigilancia*. Ministerio de Salud. 479, semana epidemiológica 52/2019. ISSN 2422-698X. Disponible en: https://www.argentina.gob.ar/sites/default/files/biv_479.pdf.
- Dirección Nacional de Epidemiología y Análisis de la Situación de Salud (DNE) (2020). *Boletín Integrado de Vigilancia - edición corregida*. Ministerio de Salud. 481, semana epidemiológica 02/2020. ISSN 2422-698X. Disponible en: https://www.argentina.gob.ar/sites/default/files/biv_479.pdf.
- Foley, D. A., Phuong, L. K., Peplinski, J., Lim, S. M., Lee, W. H., Farhat, A., ... Bowen, A. C. (2021). Examining the interseasonal resurgence of respiratory syncytial virus in Western Australia. *Archives of Disease in Childhood*, 107(3), e1.2-e7.
- Li, Y., Johnson, E. K., Shi, T., Campbell, H., Chaves, S. S., Commaille-Chapus, C., ... Nair, H. (2020). *National burden estimates of hospitalisations for acute lower respiratory infections due to respiratory syncytial virus in young children in 2019 among 58 countries: a modelling study*. The Lancet Respiratory Medicine, 9(2), 175–185.
- Lucion, M. F. & Gentile, A. (2023) *Prevención de las infecciones por virus sincicial respiratorio: un futuro promisorio*. Archivos Argentinos de Pediatría;121(5):e202310098.
- Nakajo, K., & Nishiura, H. (2023). Age-specific hospitalization risk of primary and secondary respiratory syncytial virus infection among young children. *Journal of Infection*, 87(1), 50-58.

MODELANDO LA RESPUESTA DEL CULTIVO DE MAÍZ A LA APLICACIÓN DE INSUMOS POR ZONAS DE MANEJO EN ENSAYOS A ESCALA DE LOTE

CARLOS AGUSTIN ALESSO^{1,2}, PATRICIA MELINA ACETTA², MAXIMILIANO SABINO³, GONZALO PARRA³,
PABLO ARIEL CIPRIOTTI⁴

¹*CiAgro Litoral, Universidad Nacional del Litoral-CONICET*

²*Facultad de Ciencias Agrarias, Universidad Nacional del Litoral*

³*Departamento de Desarrollo, Grupo Don Mario (GDM)*

⁴*Facultad de Agronomía – IFEVA, Universidad de Buenos Aires / CONICET*

calesso@fca.unl.edu.ar

RESUMEN

La agricultura de precisión (AP) asume la interacción entre la respuesta del cultivo a las prácticas de manejo y la heterogeneidad de ambientes o zonas intra-lote. Los ensayos a escala de lote pueden ayudar a verificar dicha hipótesis y establecer recomendaciones de manejo sitio-específicas. Las características de los datos generados en estos ensayos requieren técnicas estadísticas avanzadas para lidiar con estructuras de correlación espacial dentro de las unidades experimentales. Nuestro objetivo fue explorar alternativas de análisis de ensayos a escala de lote conducidos con tecnologías de AP. Se analizaron 10 ensayos de respuesta del cultivo de maíz a la densidad de siembra (DS) y fertilización nitrogenada (N) en el sur de Córdoba y oeste de Buenos Aires. Todos los lotes se sembraron con el mismo híbrido y contaban con zonas de manejo (ZM) de alta y baja productividad. Los tratamientos se dispusieron de manera sistemática y sin repetición en sentido perpendicular al mayor gradiente de variación. Se ajustaron 8 modelos mixtos para cada ensayo considerando ZM, N, modelo cuadrático de DS como factores cruzados fijos, y distintas estructuras de efectos aleatorios para modelar la correlación espacial con distinto nivel de detalle, desde sólo considerar la parcela del ensayo hasta tener en cuenta las coordenadas planares del registro del monitor de rendimiento a distintos niveles (parcela o ensayo). En 8 de los 10 ensayos la inclusión de la correlación espacial, en función del orden en la parcela o de las coordenadas a nivel de ensayo o por parcela, removió el efecto de la respuesta a insumos por ZM que fuera observada en los modelos no espaciales. Sólo en un caso la recomendación de DS varió marginalmente según N y ZM. Claramente, la especificación incorrecta de los efectos aleatorios omitiendo la estructura de correlación espacial tiene alto impacto en las recomendaciones.

Palabras clave: **agricultura por ambientes, correlación espacial, respuesta sitio-específica, geoestadística, modelos mixtos**

INTRODUCCIÓN

La respuesta del cultivo a la aplicación de insumos o prácticas de manejo es el resultado de la interacción entre éste y las características específicas del sitio donde crece y se desarrolla (Van Ittersum & Rabbinge, 1997). La agricultura de precisión (AP) asume esta interacción e intenta ajustar las prácticas de manejo a las condiciones sitio-específicas. Los ensayos a escala de lote pueden ayudar a verificar dicha hipótesis y establecer recomendaciones de manejo sitio-específicas (Piepho et al., 2011). Las características de los datos generados en estos ensayos requieren técnicas estadísticas avanzadas para lidiar con estructuras de correlación espacial dentro de las unidades experimentales (Whelan & McBratney, 2000). Nuestro objetivo fue explorar alternativas de análisis de ensayos a escala de lote conducidos con tecnologías de AP para detectar la interacción entre prácticas de manejo y aplicación de insumos a la variabilidad intra lote.

METODOLOGÍA

Se analizaron 10 ensayos de respuesta del cultivo de maíz a la densidad de siembra (DS) y fertilización nitrogenada (N) en el sur de Córdoba y oeste de Buenos Aires. Los tratamientos de N consistieron en tres niveles de N disponible: (i) N disponible del lote (testigo), (ii) 150 kg/ha de N disponible, y (iii) 270 kg/ha de N disponible. Los tratamientos de DS consistieron en cuatro densidades objetivo: (i) 35000, (ii) 60000, (iii) 80000, y (iv) 110000 pl/ha. Todos los lotes se sembraron con el mismo híbrido (DM2773VT3P) y contaban con zonas de manejo (ZM) de alta y baja productividad. La delimitación de las zonas de manejo se realizó utilizando el índice de vegetación NDVI a partir de imágenes Sentinel para al menos tres campañas agrícolas, con 2-3 imágenes por campaña mayormente durante floración (Diciembre-Enero). Los tratamientos, combinaciones de N y DS, se dispusieron de manera sistemática y sin repetición en sentido perpendicular al mayor gradiente de variación. Los datos se sometieron a un análisis exploratorio y se analizaron individualmente a nivel ensayo. La respuesta del cultivo en cada ensayo se modeló ajustando 8 modelos mixtos considerando los efectos fijos cruzados de ZM, N, y DS como modelo cuadrático, y distintas estructuras de efectos aleatorios para modelar la correlación espacial con distinto nivel de detalle:

- A. efecto aleatorio de la parcela sobre la ordenada al origen
- B. estructura de correlación de residuos en función de la posición en el ensayo (coordenadas X e Y)
- C. efecto aleatorio de la parcela sobre la ordenada al origen y el orden de la observación dentro de la parcela.
- D. estructura de correlación de residuos por parcela en función del número de orden del punto del monitor de rendimiento dentro de la parcela
- E. estructura de correlación de residuos por parcela en función de la posición dentro de la parcela
- F. combinación de estructuras A y E
- G. efecto aleatorio de la parcela sobre la ordenada al origen y el efecto de las coordenadas X e Y.
- H. combinación estructuras A y B

En todos los casos donde se incorporó estructura de covarianza espacial, la misma fue modelada como un proceso omnidireccional esférico. La estimación de los modelos se llevó a cabo utilizando el programa estadístico R (R Core Team, 2024) y las funciones *lme* (A, C, F, G, H) y *gls* (B, D, E) del

paquete *nlme* (Pinheiro et al., 2023). La significancia de los términos fijos se evaluó mediante ANOVA Tipo III.

RESULTADOS Y DISCUSIÓN

Los datos sugieren variación de ensayo a ensayo en la respuesta del rendimiento a la variación de la densidad de siembra y las estrategias de fertilización por zonas (Fig. 1). En general la zona de baja productividad mostró los rendimientos más bajos. No obstante, en algunos casos la respuesta (o ausencia de ella) fue similar entre zonas o niveles de fertilización, mientras que en otros casos se observaron diferencias no solo de magnitud si no de función de respuesta. Esto soportaría la idea de que la validez del enfoque de AP está muy restringida a las características propias de cada ambiente y lote (Whelan & McBratney, 2000).

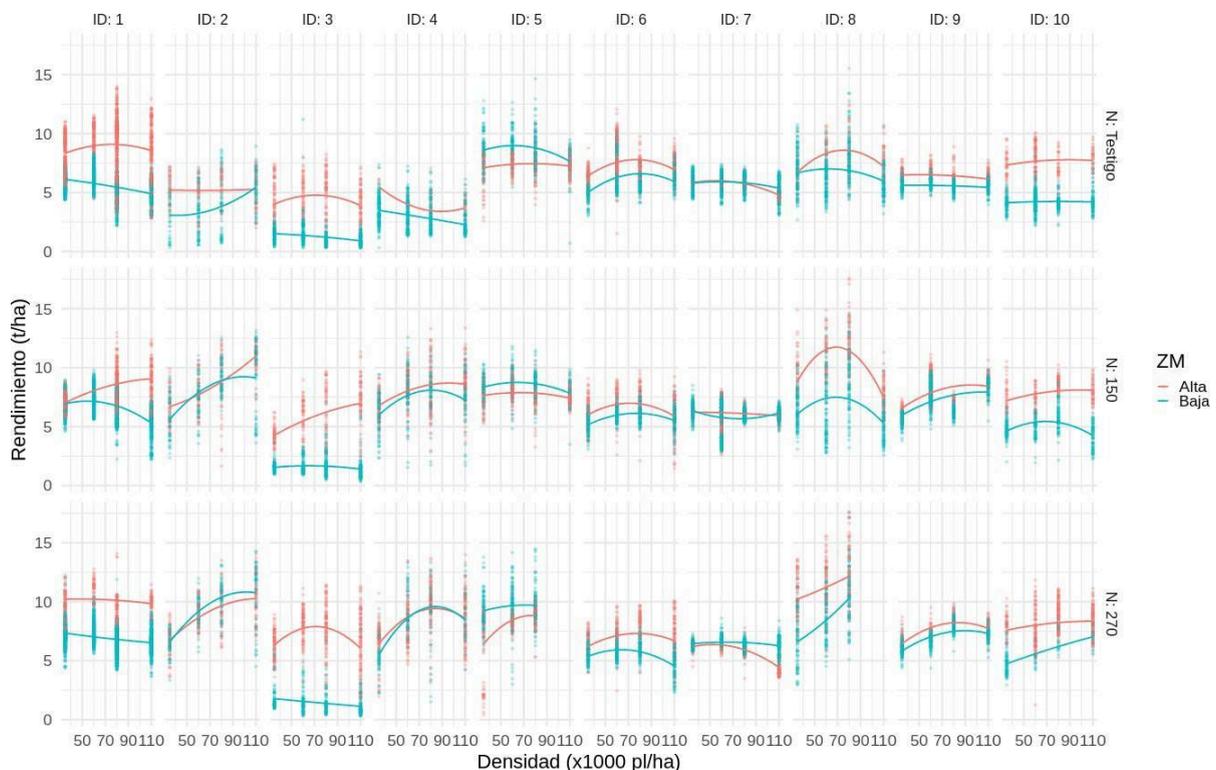


Figura 1. Patrones de respuesta de rendimiento (t/ha) a la variación de la densidad de siembra (pl/ha) y estrategias de fertilización (N) en las distintas zonas (ZM) en los 10 ensayos evaluados (ID).

Estos efectos no se convalidaron con los resultados del ANOVA del mejor modelo para cada ensayo (Tabla 1). En 8 de los 10 ensayos la inclusión de la correlación espacial, en función del orden en la parcela o de las coordenadas a nivel de ensayo o por parcela, removió el efecto de la respuesta a insumos por ZM que fuera observada en los modelos no espaciales. Sólo en un caso la recomendación de DS varió marginalmente según N y ZM. Claramente, la especificación incorrecta de los efectos aleatorios omitiendo la estructura de correlación espacial tiene alto impacto en las recomendaciones ya que en 8 de 10 ensayos la mejor recomendación sería el manejo uniforme.

Ensayo	1	2	3	4	5	6	7	8	9	10
Modelo (*)	D	E	D	E	D	B	E	E	E	E
Término	Valores p									
ZM	0.47	0.81	0.72	0.95	0.73	0.53	0.79	0.82	0.06	0.01
N	0.7	0.92	0.84	0.42	0.1	0.07	0.71	0.89	0.76	0.77
DS	0.08	0.33	0.95	0.06	0.38	0.02	0.45	0.37	0.14	0.61
DS2	0.06	0.64	0.96	0.1		0.01	0.49	0.37	0.21	0.64
ZM:N	0.93	0.86	0.05	0.76	0.17	0.44	0.56	0.82	0.88	0.04
ZM:DS	0.37	0.91	0.72	0.82	0.6	0.34	0.73	0.91	0.48	0.04
ZM:DS2	0.43	0.96	0.61	0.88		0.23	0.66	0.85	0.5	0.05
N:DS	0.91	0.56	0.96	0.06	0.03	0.01	0.66	0.85	0.69	0.68
N:DS2	0.91	0.76	0.93	0.14		0.16	0.61	0.84	0.8	0.79
ZM:N:DS	0.74	0.78	0.12	0.76	0.22	0.43	0.35	0.76	0.88	0.04
ZM:N:DS2	0.62	0.65	0.19	0.83		0.38	0.3	0.74	0.93	0.03

Tabla 1. Resumen de las tablas de ANOVA tipo III de los mejores modelos (*) ajustados para cada ensayo.

Si bien este resultado ya ha sido reportado extensivamente en la bibliografía (Alesso et al., 2019; Griffin et al., 2020), los resultados aquí expuestos tienen alta implicancia fuera del ámbito académico ya que en los últimos años, gracias a la mayor disponibilidad de maquinaria equipada con tecnologías de AP, este tipo de ensayos se han vuelto rutinarios en la industria. No obstante, el tratamiento de la correlación espacial requiere formación estadística avanzada y adaptar el análisis a las características del diseño y modo de generación de los datos. La sobresimplificación del análisis en algunos casos puede confundir los resultados obtenidos, generando recomendaciones a partir de efectos pobremente estimados. Finalmente, según lo observado en algunos ensayos, un aspecto a considerar es la pertinencia del manejo por zonas, o bien los criterios utilizados para detectar y delimitar las zonas deben ser revisados, dado que en algunos casos no se observaron diferencias sustanciales entre ellas.

REFERENCIAS

- Alesso, C. A., Cipriotti, P. A., Bollero, G. A., & Martin, N. F. (2019). Experimental Designs and Estimation Methods for On-farm Research: A Simulation Study of Corn Yields at Field Scale. *Agronomy Journal*, 111(6), 2724–2735. <https://doi.org/10.2134/agronj2019.03.0142>
- Griffin, T. W., Fitzgerald, G. J., Lowenberg-DeBoer, J., & Barnes, E. M. (2020). Modeling local and global spatial correlation in field-scale experiments. *Agronomy Journal*, 112(4), 2708–2721. <https://doi.org/10.1002/agj2.20266>



-
- Piepho, H. P., Richter, C., Spilke, J., Hartung, K., Kunick, A., & Thöle, H. (2011). Statistical aspects of on-farm experimentation. *Crop and Pasture Science*, 62(9), 721. <https://doi.org/10.1071/CP11175>
- Pinheiro, J. C., Bates, D. M., & R Core Team. (2023). *nlme: Linear and Nonlinear Mixed Effects Models* (Version 3.1) [Computer software]. R Foundation for Statistical Computing.
- R Core Team. (2024). *R: A Language and Environment for Statistical Computing* (Version 4.4) [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Van Ittersum, M. K., & Rabbinge, R. (1997). Concepts in production ecology for analysis and quantification of agricultural input-output combinations. *Field Crops Research*, 52(3), 197–208.
- Whelan, B. M., & McBratney, A. B. (2000). The “Null Hypothesis” of Precision Agriculture Management. *Precision Agriculture*, 2, 265–279

ANÁLISIS DE PATRONES DE PUNTOS PARA DETECTAR ASOCIACIONES ESPACIALES ENTRE HORMIGAS DE DESIERTO Y CARACTERÍSTICAS DE SU ENTORNO

ANA L. CAO¹, FERNANDO A. MILESI², RODRIGO G. POL^{1,3}, LUCÍA VULLO¹,
M. FLORENCIA MIRETTI¹, JAVIER LOPEZ de CASENAVE¹

¹*Ecodes, Instituto de Ecología, Genética y Evolución (IEGEB), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires-CONICET*

²*INIBIOMA (UNCo-CONICET) - Sede Junín de los Andes*

³*IADIZA (CCT-CONICET) y FCEN-UN Cuyo*

annylau2612@gmail.com

RESUMEN

La estadística espacial exploratoria y el ajuste de parámetros de modelos estocásticos a patrones de puntos permiten detectar y sopesar la importancia del componente espacial de diferentes procesos ecológicos. Buscamos y determinamos la ubicación espacial de colonias de hormigas de los géneros *Pheidole* y *Pogonomyrmex* en parcelas (2D) y en transectas (1D) en varias temporadas en la porción central del desierto del Monte (Reserva de la Biosfera de Ñacuñán, Mendoza). Estudiamos las propiedades de primer (intensidad) y segundo orden (autocorrelación espacial) de los procesos espaciales que podrían haber generado los patrones observados, poniendo a prueba los supuestos más parsimoniosos: homogeneidad de la intensidad (pruebas Kolmogorov-Smirnov de bondad de ajuste, AIC de modelos Poisson con y sin coordenadas como predictores) y ausencia de autocorrelación espacial (funciones de Ripley y de correlación de a pares con simulaciones Montecarlo). En función de esos resultados exploramos la relación de la ubicación de las colonias con variables ambientales potencialmente claves. Con análisis 1D (transectas como redes lineales de un segmento) concluimos que el agregamiento intraespecífico y la segregación interespecífica de las tres especies de *Pogonomyrmex* simpátricas se debe a sus relaciones contrastantes con las variaciones de la permeabilidad del suelo y la topografía (modelos Poisson de intensidad heterogénea con esas variables predictoras), que persisten entre años, y no con interacciones positivas/negativas entre sus colonias. Con análisis 2D (comparación de modelos Poisson y modelos Gibbs con interacción positiva/negativa entre puntos) explicamos la disposición de colonias de *Pheidole bergi* en función de la distancia a los bordes entre parches con y sin vegetación leñosa y a la intensidad de las interacciones entre colonias, que cambiaron entre años de alta y baja densidad. Con estas técnicas logramos tanto detectar como descartar factores bióticos y abióticos relevantes para la nidificación y la supervivencia de estas hormigas de desierto.

Palabras clave: *estadística espacial, spatstat, patrones de puntos, hormigas*

INTRODUCCIÓN

La estadística espacial exploratoria y el ajuste de parámetros de modelos estocásticos a patrones de puntos permiten detectar y sopesar la importancia del componente espacial de procesos ecológicos (Baddeley et al. 2016). La heterogeneidad de las características ambientales abióticas y bióticas a diferentes escalas es clave en el funcionamiento de los desiertos (Aguar & Sala 1999). La disposición espacial de las colonias de hormigas puede deberse a diferentes procesos, como la reproducción y selección del sitio de nidificación o su crecimiento y supervivencia posterior que, a su vez, pueden estar asociados con diferentes factores naturales y antrópicos (Hölldobler & Wilson 1990). En este trabajo aplicamos herramientas de la estadística espacial de patrones de puntos en una (1D) y dos (2D) dimensiones para describir e intentar comprender la asociación de las colonias de especies de hormigas de la porción central del desierto del Monte con la heterogeneidad de las características edáficas y de la vegetación leñosa en su entorno.

METODOLOGÍA

Para el análisis 2D, determinamos la ubicación espacial de las colonias de *Pheidole bergi* en 2001 (alta densidad) y 2019 (baja densidad) en dos parcelas permanentes de 50x50 m en el algarrobal de la Reserva de la Biosfera de Ñacuñán (Mendoza), en las cuales mapeamos la disposición de la vegetación en parches de plantas leñosas (> 1m altura) o sin vegetación leñosa con precisión de 1 m². Para el análisis 1D, determinamos la posición de las colonias de tres especies del género *Pogonomyrmex* (PI: *Pogonomyrmex inermis*; PM: *P. mendozanus*; PP: *P. propinqua*) en 2004 y 2007, y en 2010 después de una sequía, en dos transectas de 1500 m sobre caminos rurales que cruzan el algarrobal, y estimamos índices de permeabilidad del suelo y la posición topográfica relativa a partir de muestras de tiempo de infiltración y un relevamiento altimétrico, respectivamente.

Primero exploramos la intensidad de tipo Poisson que podrían haber generado las disposiciones de colonias y su asociación con las variables ambientales medidas. En 1D, evaluamos con pruebas de bondad de ajuste Kolmogorov-Smirnov si la intensidad varió más a lo largo de las transectas que lo esperado por modelos nulos de intensidad homogénea. Y en 1D y 2D, determinamos con pruebas de desviación máxima (MAD) si la autocorrelación espacial estimada con funciones de Ripley y de correlación de pares excedió los sobres de confianza de simulaciones Montecarlo de un proceso espacial tipo Poisson completamente aleatorizado (CSR) (Baddeley et al. 2014, 2016). Luego, integrando todos los patrones observados y considerando a las transectas como réplicas, modelamos a la intensidad lineal según modelos Poisson heterogéneos con funciones cuadráticas interactuantes de las dos variables edáficas y parámetros potencialmente cambiantes entre años y especies. En 2D, con las parcelas como réplicas, ajustamos modelos Gibbs (los residuos de modelos Poisson aún mostraban agregamiento) con la intensidad (primer orden) como función del año y del tipo de parche o la distancia a los bordes de la vegetación leñosa (estableciendo distancias negativas para píxeles dentro de los parches con vegetación) y la autocorrelación (segundo orden) con funciones Strauss-Hard de magnitud y radio potencialmente variables para colonias de la misma o de diferente especie. En todos los casos evaluamos la importancia de los predictores comparando la devianza de modelos alternativos con el criterio AIC de Akaike (AIC).

Los análisis fueron realizados usando *spatstat* (versión 3.0-7; Baddeley et al. 2016) en la plataforma R

(R Core Team 2022; version 4.0.4).

RESULTADOS Y DISCUSIÓN

Las tres especies de *Pogonomyrmex* mostraron densidades heterogéneas todos los años y en mayor o menor grado a lo largo de ambas transectas (Fig. 1).

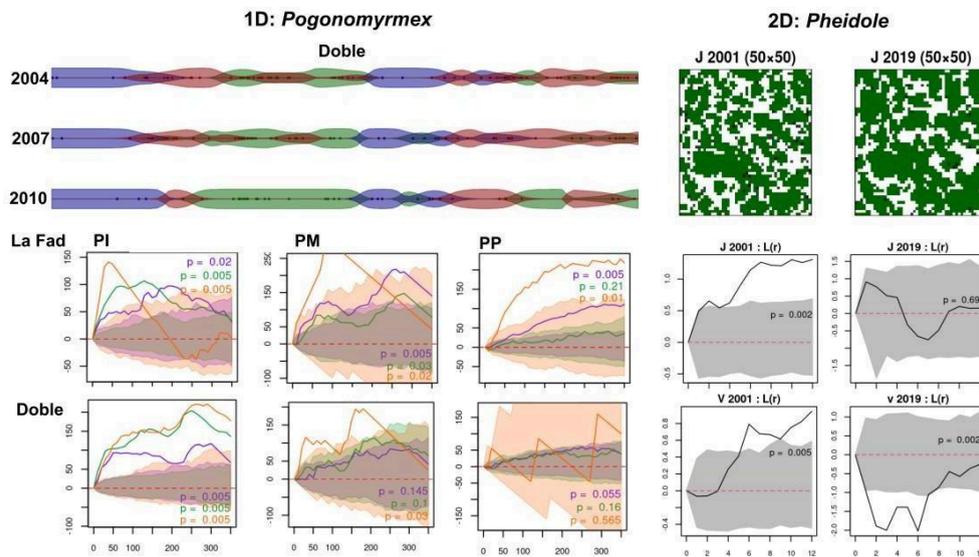


Fig. 1. Arriba: 1D: colonias de cada una de las tres especies de *Pogonomyrmex* (PI: verde, PM: azul, PP: rojo) a lo largo de una de las transectas en los tres años muestreados con bandas representando la probabilidad variable de que una colonia pertenezca a cada especie; 2D: mapa de parches de vegetación leñosa en una de las parcelas, con las colonias como pixeles negros. Abajo: 1D: funciones K de Ripley (centradas) observadas en función de la distancia r entre colonias (2004: violeta, 2007: verde, 2010: naranja), con los sobres de confianza del 95% para modelos CSR (valor esperado = 0) y los p valores que resultan de pruebas globales MAD; 2D: idem con funciones L de Ripley (versión centrada y estandarizada de K).

El modelo Poisson seleccionado mostró una estrecha relación de la densidad de dos de las especies (PI y PM) con valores opuestos de permeabilidad del suelo y de posición topográfica (altura relativa), explicando su segregación espacial, y de la tercera (PP) con los valores intermedios más típicos de las condiciones del algarrobal (Fig. 2, Tabla 1). Esas relaciones no cambiaron con el tiempo, por lo que los sectores con colonias de cada especie se mantuvieron aún cuando variaron notoriamente sus densidades (como en 2010, luego de una sequía). Tanto el agregamiento intraespecífico como la segregación interespecífica de las tres especies de *Pogonomyrmex* simpátricas se puede explicar como consecuencia de sus relaciones contrastantes con las variaciones edáficas ligadas a la dinámica del agua de lluvia (escorrentía, acumulación, infiltración y persistencia), sin necesidad de involucrar interacciones positivas/negativas directas entre sus colonias.

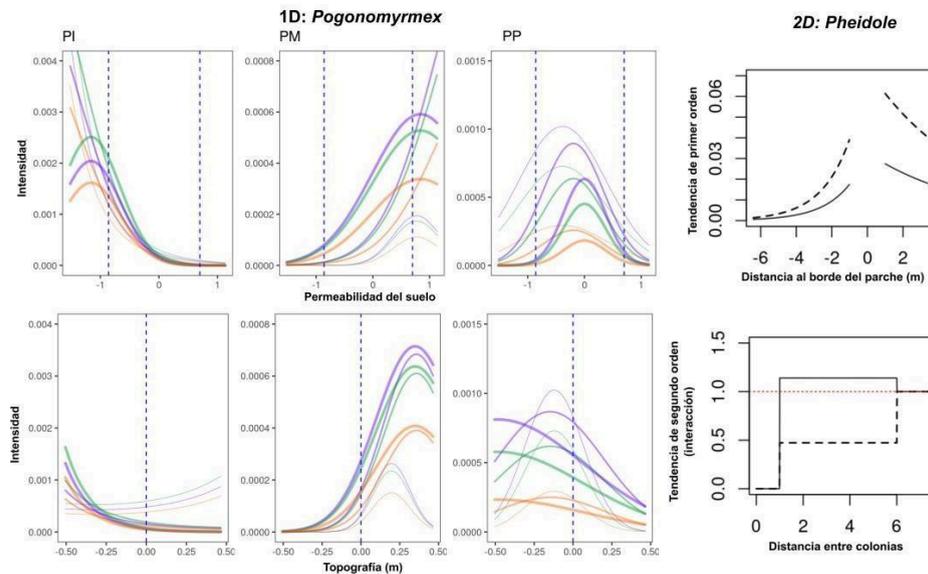


Fig. 2. 1D: Efectos parciales de los índices de permeabilidad del suelo (arriba) y de posición topográfica (abajo) sobre la intensidad del proceso espacial que resulta en la ubicación de las colonias (2004: violeta, 2007: verde, 2010: naranja) para los tres cuartiles de la otra variable en orden de grosor de línea, según el modelo Poisson elegido (“Interacción”, Tabla 1). Las líneas verticales marcan umbrales críticos (índices de permeabilidad entre ambientes típicos del sitio de estudio, y altura igual al entorno). 2D: Efecto de la distancia al borde entre parches (negativas hacia dentro del parche de arbustos) sobre la intensidad condicional (arriba) y de la distancia entre colonias sobre la autocorrelación (abajo) según el modelo Strauss-Hard elegido (“BordeHet+”, Tabla 1) que resulta en la ubicación de las colonias en 2001 (contínuas) y 2019 (discontinuas).

Las colonias de *Pheidole bergi* presentaron una disposición espacial agrupada cuando su densidad fue alta (2001) pero uniforme o compatible con lo esperado por CSR cuando su densidad se redujo a la mitad (2019; Fig. 1). Entre las alternativas ensayadas de relación de la disposición de colonias con la estructura de parches de vegetación, el mejor modelo fue el que tiene la mayor probabilidad de colonias en los bordes entre parches y que desde allí disminuye con diferente pendiente según el tipo de parche (Tabla 1). La interacción adicional entre pares de colonias vecinas (1–6 m de distancia) fue ligeramente positiva en densidades altas, resultando en una agregación extra de las colonias, y fuertemente negativa en densidades bajas, resultando en un patrón más uniforme (Fig. 2). Si bien ambos componentes están parcialmente confundidos (porque están asociados a una misma escala espacial), la disposición de colonias de *P. bergi* se puede explicar en relación a la vegetación leñosa y a una repulsión entre colonias cercanas.

1D: <i>Pogonomyrmex</i>			2D: <i>Pheidole</i>		
Modelo	Estructura	AIC	Modelo	Estructura	AIC
Nulo	spp	6234.73	CSR	año	2231.25
Año	spp x año	6210.76	HOM+	año	2166.97
Camino	spp x año x camino	6220.82	HET	año + veg/sd	2209.3
SPI 2	spp x año + spp x SPI ²	6002.46	HET+	año + veg/sd	2142.3
Independiente	spp x año + spp x (SPI ² + TPI ²)	5971.3	Borde	año + abs(dist)	2216.34
Interacción	spp x año + spp x SPI² x TPI²	5937.87	Borde+	año + abs(dist)	2144.63
Full	spp x año x SPI ² x TPI ²	5999.7	Borde Het	año + abs(dist) * veg/sd	2200.03
			Borde Het+	año + abs(dist) * veg/sd	2126.89
			Distancia ²	año + dist ²	2202.6
			Distancia ² +	año + dist ²	2130.04

Tabla 1: Modelos espaciales Poisson ensayados para los patrones de puntos dados por colonias activas de tres especies de *Pogonomyrmex* sobre caminos (1D) y modelos Poisson y Gibbs (con interacción Strauss-Hard, indicados con “+”) para las colonias activas de *Pheidole bergi* en parcelas de algarrobal (2D), con las estructuras de variables predictoras y los valores AIC correspondientes. Los modelos elegidos se destacan en negritas.

Mediante la estadística espacial descriptiva y el modelado de procesos espaciales de patrones de puntos en una y dos dimensiones logramos tanto detectar como descartar factores bióticos y abióticos, importantes para la nidificación, supervivencia y coexistencia de hormigas de desierto.

REFERENCIAS

- Aguiar, M. R., & Sala O. E. (1999). Patch structure, dynamics and implications for the functioning of arid ecosystems. *Trends in Ecology & Evolution*, 14, 273-277.
- Baddeley, A., Diggle, P.J., Hardegen, A., Lawrence, T., Milne, R.K., & Nair G. (2014). On tests of spatial pattern based on simulation envelopes. *Ecological Monographs*, 84, 477-489.
- Baddeley, A., Rubak E. & Turner R. (2016). *Spatial point patterns, methodology and applications with R*. Boca Ratón: CRS Press.
- Baddeley, A., Turner R., & Rubak, E. (2016). *Spatstat: Spatial point pattern analysis, model fitting, simulation and test (v3.0-7) [R package]* [spatstat - spatstat website](https://www.r-project.org/web/packages/spatstat/vignettes/spatstat.html).
- Hölldobler, B., & Wilson, E.O. (1990). *The ants*. Cambridge: Harvard University Press.
- R Core Team (2022) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Viena.

ESTIMACIÓN DE CORRELACIONES ENTRE RENDIMIENTOS DE CULTIVOS A PARTIR DE REGISTROS EMPRESARIALES

LUIS FRANK¹, FACUNDO LANDAJÓ², CRISTINA RAS²

¹Depto. Métodos Cuantitativos y Sistemas de Información, Facultad de Agronomía, Universidad de Buenos Aires

²Depto. de Economía, Desarrollo y Planeamiento Agrícola, Facultad de Agronomía - Universidad de Buenos Aires

frank@agro.uba.ar

RESUMEN

Se propone una metodología de estimación de coeficientes de correlación entre rendimientos de cultivos a nivel de empresa. El método se basa en la combinación de una regresión lineal y su versión revertida en un único modelo, uno de cuyos parámetros es el coeficiente de interés. El método resulta particularmente útil para estimar correlaciones en situaciones donde los tamaños muestrales son extremadamente pequeños, como ocurre cuando los planes de producción de la empresa cambian año tras año, y ciertos cultivos rara vez se producen juntos. A lo largo del artículo se realiza un ejercicio de estimación con datos reales, y los resultados se comparan con estimaciones previas calculadas a partir de estadísticas públicas. Se concluye que las correlaciones calculadas a nivel de empresa son en general mayores que las calculadas a nivel agregado, al menos para el área de estudio. Asimismo, se halla evidencia de que las correlaciones entre rendimientos no son fijas para el mismo sitio o empresa, sino que pueden variar según la fase del ciclo climático ENSO, al menos para ciertos pares de cultivos. Sobre la base de estos resultados, se proponen algunas guías para construir matrices de correlaciones, cuando no se disponga de datos a nivel de empresa.

Palabras clave: *simulación, agricultura, administración rural.*

INTRODUCCIÓN

Simular el resultado económico de una empresa agrícola requiere una estimación previa de la matriz de correlaciones entre rendimientos de aquellos cultivos que constituyen el plan de producción. El consenso profesional sugiere que dicha matriz es, o debería ser, específica para áreas productivamente homogéneas (Goodwin & Mishra, 2002), es decir, para áreas que comparten características ambientales comunes, y cuyos productores realizan prácticas culturales similares y tienen acceso a la misma tecnología. Una definición tan estricta de homogeneidad, sin embargo, dificulta la estimación de la matriz de correlaciones porque requiere bases de datos de rendimientos extremadamente grandes, provenientes preferiblemente de la misma empresa o de empresas ubicadas próximas entre sí, que abarquen períodos sin cambios tecnológicos relevantes. Como este tipo de bases de datos generalmente no están disponibles, una práctica común consiste en calcular las correlaciones a partir de series de rendimientos no demasiado largas del mismo distrito de la base de la ex Secretaría de Agricultura, Ganadería, Pesca y Alimentación de la Nación (SAGPyA), o incluso a partir de rendimientos simulados con modelos agronómicos como DSSAT. Ejemplos de correlaciones obtenidas a través de modelos agronómicos de simulación de rendimientos pueden encontrarse en

Ciudad de Buenos Aires, octubre 2024

(Westen, 2009). No obstante, se desconoce si esta aproximación produce estimaciones precisas de los verdaderos coeficientes de correlación o tienen algún tipo de sesgo.

La revisión bibliográfica y la consulta con expertos revelan que no existe una metodología uniforme para estimar correlaciones entre rendimientos a nivel de empresa, más allá de algunas recomendaciones generales. Por ejemplo, se recomienda utilizar el coeficiente de correlación de Spearman en vez del coeficiente de Pearson a fin de captar posibles asociaciones no lineales entre rendimientos. Asimismo, se recomienda utilizar una cantidad limitada de campañas agrícolas para evitar sesgos de estimación debidos a cambios tecnológicos. Estas recomendaciones, sin embargo, no resuelven el principal problema de los datos micro: que el plan de producción de la empresa no incluye los mismos cultivos todos los años, de manera que la cantidad de pares de rendimientos disponibles en un período sin cambio tecnológico es realmente escasa. El trabajo que sigue propone una estrategia de estimación para lidiar específicamente con este problema. En primera instancia expondremos el método, con su correspondiente justificación teórica, y en una segunda instancia estimaremos una matriz de correlaciones con este método a partir de datos reales.

METODOLOGÍA

Consideremos el par de rendimientos $\{X_1, X_2\}$ correspondientes a dos cultivos de una unidad productiva homogénea. La correlación entre X_1 y X_2 se mantiene constante si las condiciones tecnológicas y ambientales en que se desarrollan dichos cultivos no cambian. En períodos cortos de tiempo, por ejemplo, menos de una década, es razonable suponer que la tecnología de producción no cambia. No obstante, incluso en períodos tan cortos las condiciones ambientales pueden variar (al menos en sudamérica) en función de la fase de la Oscilación del Pacífico Sur “El Niño” o ENSO. En consecuencia, la relación entre X_1 y X_2 se puede expresar como una relación lineal que interactúa con las fases del ENSO. Formalmente, esta relación se escribe

$$X_{i1} = \beta_1 + (\beta_2 + \delta_{\text{ENSO}} \beta_3) X_{i2} + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, \dots, n$$

donde δ_{ENSO} es una variable ternaria que toma los valores $\{1, 0, -1\}$ en años húmedos (fase del Niño), neutros, o secos (fase de la Niña), respectivamente. Esta misma relación es válida para expresar la relación entre rangos de rendimientos, es decir $R_i(X_{ij})$, donde el operador $R_i(\cdot)$ es el que convierte los rendimientos del i -ésimo cultivo en rangos. Lógicamente, los coeficientes β_i del modelo transformado son distintos a sus homólogos del modelo sin transformar. El modelo de rangos puede a su vez estandarizarse, promediando las observaciones en ambos lados de la ecuación y dividiendo por los respectivos desvíos estándar, de manera que

$$Z_{i1} = \beta_2 \left(\frac{\sigma_{R(X_2)}}{\sigma_{R(X_1)}} \right) Z_{i2} + \beta_3 \left(\frac{\sigma_{R(X_3)}}{\sigma_{R(X_1)}} \right) Z_{i3} + \epsilon_i^*, \quad \text{donde } \epsilon_i^* \sim N(0, 1),$$

y la variable Z_{i3} es la versión estandarizada de $R_3(X_{i3}) = R_2(\delta_{\text{ENSO}} X_{i2})$. Ahora bien, los cocientes de desvíos estándar de la expresión anterior se cancelan porque los $\sigma_{R(X_j)}$ sólo dependen del tamaño de la muestra, de manera que β_2 y β_3 representan directamente los coeficientes de correlación de Spearman entre X_1 y X_2 , y entre X_1 y X_3 , respectivamente. Esta última afirmación no es obvia y la demostraremos por medio de la proposición que sigue. Nótese que, al transformar en rangos, se torna irrelevante estandarizar las variables X .

PROPOSICIÓN 1. En el modelo clásico de regresión, cada coeficiente representa la correlación parcial

entre el regresando y el regresor correspondiente multiplicada por el cociente de varianzas entre ellos.

Demostración. Consideremos el modelo clásico de regresión particionado

$$y = Z\beta_{k-1} + x_k\beta_k + \epsilon \quad \epsilon \sim N(0, \sigma^2 I_n).$$

Donde Z es la matriz de todos los regresores de X excepto el último, x_k . El significado de estas letras es propio de esta proposición y no se relaciona en absoluto con el significado de las mismas letras en otras partes del texto. El estimador mínimo cuadrático de β es

$$b_{OLS} = \begin{bmatrix} Z'Z & Z'x_k \\ x_k'Z & x_k'x_k \end{bmatrix}^{-1} \begin{bmatrix} Z'y \\ x_k'y \end{bmatrix}$$

Pero recordando que la inversa de una matriz A particionada en bloques puede descomponerse (Ton-Tzer Ly & Shiu Sheng-Hua, 2002) como

$$A^{-1} = \begin{bmatrix} (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1} & -(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1} \\ -(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1} & (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} \end{bmatrix}$$

si A_{11}^{-1} y A_{22}^{-1} existen, el estimador del k-ésimo coeficiente b_k es

$$b_k = \frac{-x_k'Z(Z'Z)^{-1}Z'y + x_k'y}{x_k'x_k - x_k'Z(Z'Z)^{-1}Z'x_k}, \quad \text{o bien, si } M = Z(Z'Z)^{-1}Z',$$

$$b_k = \frac{x_k'(I_{k-1} - M)y}{x_k'(I_{k-1} - M)x_k} = \frac{e_x'e_y}{e_x'e_x} = \frac{e_x'e_y}{\sqrt{e_x'e_x}} \frac{\sqrt{e_y'e_y}}{\sqrt{e_y'e_y}} \frac{1}{\sqrt{e_x'e_x}} = r_{x,y} \frac{\sqrt{e_y'e_y}}{\sqrt{e_x'e_x}}$$

Donde M es una matriz idempotente, al igual que $(I_{k-1} - M)$, y $r_{x,y}$ es un estimador del coeficiente de correlación parcial entre x_k e y. Se puede verificar fácilmente que las sumas de cuadrados del numerador y denominador son efectivamente sumas de cuadrados de los residuos, ya que

$$e_x = x_k - Z(Z'Z)^{-1}Z'x_k \quad \text{y} \quad e_y = y - Z(Z'Z)^{-1}Z'y$$

La decisión de expresar X_1 en función de X_2 en el modelo original es totalmente arbitraria y bien podríamos haber expresado X_2 en función de X_1 , en cuyo caso la relación de variables estandarizadas sería, para muestras no demasiado pequeñas,

$$Z_{i2} = \alpha_1 Z_{i1} + \alpha_3 Z_{i3}^* + \epsilon_i^*, \quad \epsilon_i^* \underset{\text{aprox.}}{\sim} N(0,1).$$

Pero si tanto β_2 como α_1 representan coeficientes de correlación (de Spearman en el modelo transformado en rangos y de Pearson en el modelo original), entonces $\beta_2 = \alpha_1$, lo cual habilita el planteo del modelo combinado

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} = \begin{bmatrix} Z_2 & Z_3 & 0_n \\ Z_1 & 0_n & Z_3 \end{bmatrix} \begin{bmatrix} \beta_2 \\ \beta_3 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \epsilon_1^* \\ \epsilon_2^* \end{bmatrix} \quad \text{donde } \epsilon^* \underset{\text{aprox.}}{\sim} N(0_{2n}, I_{2n})$$

La posibilidad de combinar cada modelo con su versión revertida es fundamental para estimar las correlaciones parciales entre X_1 y X_2 cuando los tamaños muestrales son muy reducidos, ya que la combinación duplica la cantidad de observaciones útiles. A continuación, veremos un caso de estudio.

RESULTADOS

Estimamos los coeficientes de correlación de Spearman entre pares de rendimientos de cuatro cultivos a partir de una base de datos de una empresa administradora de campos en el SE de la provincia de Buenos Aires. Los rendimientos fueron registrados entre las campañas 2016/17 y 2023/24 en 47 establecimientos administrados por la empresa. A los fines del cálculo, los cultivos de soja y maíz discriminados por época de siembra se trataron como cultivos diferentes, así como el girasol confitero y el destinado a la industria. Al seguir cada establecimiento un plan de producción propio, no todos los cultivos se repetían todos los años en los mismos establecimientos. El protocolo de cálculo fue el siguiente:

Análisis exploratorio. En primera instancia, se realizó un ANCOVA exploratorio en el que los rendimientos (variable respuesta) fueron clasificados por cultivo-época de siembra y establecimiento (siempre que tuvieran más de tres observaciones), y el tiempo se introdujo como covariable. Dado que este último no resultó significativo, se repitió el análisis reemplazando el tiempo por una variable clasificatoria de las fases del ENSO correspondientes a cada campaña siguiendo un esquema de diseño factorial desbalanceado.

Categorización y cálculo de correlaciones. En una segunda instancia, se utilizaron los tests de comparaciones múltiples (Tukey y Duncan) realizados en la etapa anterior para agrupar a los establecimientos y cultivos en unidades o categorías homogéneas. Los establecimientos se agruparon en 7 conjuntos y los cultivos en 6 categorías. Luego se promediaron los rendimientos de las $7 \times 6 \times 3 = 126$ combinaciones de unidades productivas, cultivos según época de siembra, y fases del ENSO, aunque 38 de estas combinaciones no registraron rendimiento. Finalmente, los rendimientos de cada cultivo fueron convertidos en rangos y estandarizados, y se calcularon las correlaciones entre pares a través de los coeficientes de regresión del modelo combinado.

Una vez estimados los coeficientes de correlación no nos extendimos en la verificación de los supuestos subyacentes al modelo combinado, salvo por la verificación del supuesto de independencia lineal entre los regresores, ya que es sabido que dependencias moderadas a fuertes pueden derivar en estimaciones con signo opuesto al esperado. La inspección de las correlaciones entre los regresores reveló dependencias débiles o incluso nulas. Por otra parte, los coeficientes β_3 y α_3 resultaron no significativos en la mayoría de los casos. El cuadro 1 muestra la matriz de correlaciones obtenida siguiendo la secuencia del cálculo descripta anteriormente.

	Maíz tardío	Girasol	Soja 1era	Trigo	Soja 2da	Maíz
Maíz tardío	1					
Girasol	0,9022	1				
Soja 1era	0,8015	0,8235	1			
Trigo	0,8044	0,7720	0,8514	1		
Soja 2da	0,7089	0,5587	0,5998	0,8132	1	
Maíz	0,5143	0,5117*	0,0446*	0,1099*	0,6139	1

Cuadro 1. Matriz de correlaciones entre pares de rendimientos. Los * indican valores- p superiores a 0,025 en el análisis de regresión.

La matriz del cuadro 1 se comparó con una matriz de consenso calculada por (Frank, 2024) para la región pampeana y con una matriz específica de la zona de Balcarce, ambas estimadas a partir de series de rendimientos de SAGPyA. Las comparaciones se realizaron por medio del test de Jennrich (Jennrich, 1970; Atiany, 2016) y considerando tamaños muestrales medios de 25 observaciones para la matriz del cuadro 1 y de 35 para las otras dos. El test de Jennrich se programó en Euler Math Toolbox. El código está disponible a solicitud del interesado. Utilizamos tamaños muestrales medios para realizar el test porque cada elemento de la matriz de correlaciones se calculó por separado y con una cantidad distinta de rendimientos, situación no prevista por Jennrich. Los cultivos contrastados fueron soja de 1era., soja de 2da., maíz (o maíz tardío en nuestro caso), girasol y trigo. El valor- p asociado al estadístico χ^2 de Jennrich para la comparación de nuestra matriz con la de consenso fue 0,002; en tanto que el mismo valor para la comparación con la matriz específica de Balcarce fue 0,01. Es decir, en ambos casos rechazamos la hipótesis de igualdad entre matrices.

DISCUSIÓN

El trabajo precedente propone un método de estimación de correlaciones entre rendimientos de cultivos basado en la combinación de una regresión de rangos y de su versión revertida. El método propuesto es particularmente útil para estimar correlaciones a nivel de empresa, dado que los planes de producción cambian de una campaña a otra y la cantidad de pares de rendimientos disponibles suele ser extremadamente reducido. El nuevo método se aplica al cálculo de correlaciones entre rendimientos de soja (de primera y de segunda), maíz, girasol y trigo, de una empresa real ubicada en el SE de la provincia de Buenos Aires y la matriz resultante se compara con matrices análogas, pero calculadas a partir de datos de la ex SAGPyA, una de las cuales es específica del partido de Balcarce y la otra es una matriz de consenso para toda la región pampeana (Frank, 2024).

Los resultados muestran (a) que el nuevo método arroja correlaciones (significativamente) positivas, en línea con estimaciones previas (ver e.g. Westen, 2009 y Frank, 2024) obtenidas a través de métodos y fuentes de información diversas, incluso modelos de simulación agronómica; (b) que las correlaciones a nivel de empresa son mayores que las correlaciones obtenidas a nivel de partido o incluso a nivel regional, y estas diferencias son significativas; y (c) que las correlaciones no son fijas para un determinado sitio, sino que además pueden variar en función del ciclo climático. Esta última conclusión, no obstante, debe tomarse con precaución dado que hasta el momento sólo ha sido verificada en nuestro caso de estudio y para correlaciones entre maíz tardío y demás cultivos de

verano. Los puntos (a) y (b) proveen una guía para acotar las correlaciones a nivel de empresa a los límites impuestos por el mínimo valor “significativo” de r , es decir $P(R < r) = \alpha$, y el valor estimado para la zona en cuestión. Para $\alpha = 0,05$, por ejemplo, las cotas (en escala logarítmica y fracciones simples) son

$$\max\left\{\frac{8}{9} - \frac{16}{25} \ln n; \ln r_{\text{zona}}\right\} < \ln \rho < 0$$

Donde n es la duración (años) del período sin cambio tecnológico, r_{zona} es el estimador de la correlación para el mismo par de cultivos a nivel zonal.

REFERENCIAS

- Atiany, T. A. M. & Shamshuritawati, S. (2016). New Statistical Test for Testing Several Correlation Matrices. *Global Journal of Pure and Applied Mathematics*, 12(5): 4285-4298.
- Frank, L. (2024). Coeficientes de correlación entre rendimientos de cultivos pampeanos estimados a partir de series de SAGPyA. XXVIII Reunión Científica del Grupo Argentino de Bioestadística. 8 al 10 de octubre de 2024. Buenos Aires, Argentina.
- Goodwin, B. K. & Mishra, A. K. (2002). An Empirical Evaluation Of Yield Performance And Cross-Crop Yield Correlation. *2002 Annual Meeting of the Western Agricultural Economics Association*. July 28-31, Long Beach, California. <https://ideas.repec.org>
- Jennrich, R. I. (1970). An asymptotic χ^2 test for the equality of two correlation matrices. *Journal of the American Statistical Association*, 65(330): 904-912.
- Ton-Tzer, Lu & Shiou, Sheng-Hua (2002). Inverses of 2×2 Block Matrices. *Computers and Mathematics with Applications* 43: 119-129.
- Vose, D. (2000). Risk Analysis. *A Quantitative Guide*. Second Edition. Wiley & Sons. Chichester.
- Westen, I. S. (2009). Análisis del uso de distintos coeficientes de correlación para la estimación del riesgo de un portfolio agrícola con simulación Monte Carlo. Tesis de grado. Facultad de Agronomía, UBA.



CORRELACIONES ENTRE RENDIMIENTOS DE CULTIVOS PAMPEANOS ESTIMADAS A PARTIR DE ESTADÍSTICAS PÚBLICAS

LUIS FRANK¹

¹Depto. Métodos Cuantitativos y Sistemas de Información, Facultad de Agronomía, Universidad de Buenos Aires

frank@agro.uba.ar

RESUMEN

La simulación del resultado económico de una empresa agrícola requiere la introducción de una matriz de correlaciones entre rendimientos de cultivos. No existe, sin embargo, una metodología comúnmente aceptada por los analistas de riesgo agropecuario para estimar dicha matriz, ni una matriz teórica de referencia que la sustituya en caso de que no pueda ser estimada por falta de datos. En el trabajo se presenta una guía metodológica para la estimación de matrices a nivel de partido o departamento, y se estima una matriz de consenso para toda la región pampeana a partir de rendimientos registrados por la ex SAGPyA (con discontinuidades) durante 54 campañas agrícolas en 10 distritos de la región. Los resultados ubican a las correlaciones entre soja, maíz y sorgo en el orden de 0,5; entre girasol y los demás cultivos de verano, entre 0,2 y 0,3; y entre trigo y los cultivos de verano, entre 0,1 y 0,2. La correlación entre soja de primera y soja de segunda se ubica en 0,7. Estas estimaciones constituyen, en principio, cotas inferiores para la construcción de matrices a nivel subregional o local, de acuerdo a la hipótesis que sugiere que a mayor homogeneidad productiva del área bajo estudio, mayor la correlación entre rendimientos. Sin embargo, la comparación de la matriz de consenso con matrices empíricas de diverso origen (registros empresariales o modelos de simulación agronómica) no validan esta afirmación en general.

Palabras clave: **simulación, agricultura.**

INTRODUCCIÓN

El análisis de riesgo de la producción agrícola se basa con frecuencia en la simulación de resultados económicos de una empresa representativa sometida a *shocks* de distinta índole. Por medio de la simulación se procura obtener, por ejemplo, la distribución empírica de los resultados de la empresa a través de la generación de rendimientos aleatorios de aquellos cultivos que integran el plan de producción. Es decir, se trata de un procedimiento computacional que devuelve una sucesión de resultados económicos b_1, \dots, b_n a partir de una sucesión de vectores de rendimientos y_1, \dots, y_n aleatorios, cuya distribución de probabilidad se conoce de antemano. Dichos resultados se pueden resumir luego en una distribución de frecuencias que represente razonablemente la verdadera distribución de probabilidad del resultado económico. La siguiente función,

$$b_i = p'y_i - \sum_{j=1}^m w'x_j, \quad \text{donde } y_i \sim N(\mu, \text{var}(y))$$

representa el margen bruto (MB) de una empresa en la que la variabilidad del MB depende únicamente de la variabilidad del rendimiento de los m cultivos incluidos en el plan de producción. En la práctica, la distribución de las variables aleatorias que intervienen en la simulación suele ser desconocida y es el analista quien propone una forma funcional, por ejemplo, $y_i = m + Az_i$ donde $z_i \sim N(0,1)$, en cuyo caso $\text{var}(y) = AA'$ (Lütkepohl 1996, pág. 157-158). Los programas informáticos suelen requerir, en vez de $\text{var}(y)$, estimaciones separadas de las varianzas de cada variable y de las correlaciones entre ellas, y a partir de éstas construyen A (ver e.g. Vose, 2000) apelando a la relación $R = D^{-1/2}AA'D^{-1/2}$ donde D es una matriz diagonal de varianzas y R una matriz de correlaciones. La revisión bibliográfica revela que no existe una metodología uniforme para estimar R , ni valores de referencias para la región pampeana que puedan combinarse con información muestral específica de cada zona o subregión. El objetivo de este trabajo es justamente estimar la matriz de correlaciones de rendimientos R de los cinco principales cultivos de la región pampeana. Para ello, utilizaremos series históricas de rendimientos a nivel de partido o departamento del Sistema de Estimaciones Agrícolas y propondremos una metodología de estimación reproducible por cualquier analista, la que conducirá a una matriz de *consenso* para toda la región pampeana.

METODOLOGÍA

Selección de la muestra y cálculo de correlaciones:

Se seleccionaron 10 distritos representativos de la región pampeana. La selección se basó en un estudio previo inédito que clasificó los partidos o departamentos de las provincias pampeanas de acuerdo al perfil productivo que surge del Censo Nacional Agropecuario de 2018. Los partidos seleccionados para este estudio fueron Pergamino, Trenque Lauquen, Gral. Villegas, Castelli, Balcarce y Puán, en la provincia de Buenos Aires; y los departamentos de Maraco (Gral. Pico), Castellanos (Rafaela), Río Cuarto, y Villaguay en las provincias de La Pampa, Santa Fe, Córdoba y Entre Ríos, respectivamente.

Se compilaron series de rendimientos de trigo, maíz, soja, girasol y sorgo de los mencionados distritos. Las series provinieron de la base de datos de la Secretaría de Agricultura, Ganadería y Pesca de la Nación (<https://datosestimaciones.magyp.gob.ar/>) y abarcaron las campañas 1969/70 a 2022/23.

A continuación, se ajustaron sendas regresiones lineales a fin de descontar a cada serie los rendimientos que cabía esperar por el simple transcurso del tiempo. El objetivo de esta "corrección" fue descontar de cada serie el efecto del avance genético y tecnológico. En algunos casos, los coeficientes de regresión fueron estimados a través del procedimiento iterativo de Cochran Orcutt (Gujarati & Porter, 2009 pág. 455-456), pero con una sola iteración ya que es sabido (Greene, 2008 sec. 19.9, pág. 648-649) que la iteración no produce mayor ganancia de eficiencia. Una vez corregidos los coeficientes de regresión se calcularon nuevamente los residuos.

A partir de los residuos del ajuste lineal se calcularon los coeficientes de correlación de Pearson y de Spearman entre los 21 pares de cultivos. La significatividad de las correlaciones de Pearson y Spearman fueron puestas a prueba mediante diferentes *tests*. En el primer caso, se calculó un estadístico $t_{ij} = r_{ij} [(n-2)/(1-r_{ij}^2)]^{1/2}$, el que se comparó con el estadístico crítico $t_{(0,975; n-2)}$. En el segundo caso, el de los coeficientes de Spearman, las correlaciones fueron transformadas en una variable normalizada, el estadístico Z de Fisher (1934, pág. 183-192), cuya fórmula es

$$Z_{ij} = \frac{1}{2} \ln \frac{1 + r_{ij}}{1 - r_{ij}}.$$

El estadístico de Fisher sigue una distribución aproximadamente normal con esperanza nula y desvío estándar $(n-3)^{-1/2}$, de manera que el valor crítico para pruebas de significatividad al 5% es $1,96/(n-3)^{1/2}$. Los resultados de las pruebas *t* sobre coeficientes de correlación de Pearson y Z sobre coeficientes de Spearman fueron en general coincidentes.

Pruebas de hipótesis

Al cabo de esta secuencia de cálculos se dispuso de dos matrices de correlaciones entre pares de cultivos por cada distrito. La variabilidad del coeficiente de correlación entre zonas para el mismo par de cultivos fue notable, así como la superposición de rangos entre pares. Para probar diferencias entre pares de cultivos, se construyeron dos conjuntos de datos: uno con todas las combinaciones de pares de trigo, soja total, maíz, girasol y sorgo; y otro con las combinaciones de trigo, soja de 1era., soja de 2da., maíz, girasol y sorgo. Cada conjunto se asimiló a un diseño experimental en bloques completos al azar (DBCA) en el que las zonas son los bloques y los coeficientes de Spearman transformados, los tratamientos. Para mantener el DBCA balanceado algunos datos faltantes fueron reemplazados por estimaciones realizadas con el conocido procedimiento iterativo de Yates (Yates, 1933), quien sugiere reemplazar los datos faltantes por

$$\hat{z}_{ij} = \frac{b z_{i.} + t z_{.j} - z_{..}}{(t-1)(b-1)}, \quad i = 1, \dots, b; j = 1, \dots, t$$

Donde $z_{i.}$ es la suma de datos observados del *i*-ésimo bloque, $z_{.j}$ es la suma de datos observados del *j*-ésimo tratamiento, y $z_{..}$ es la suma total de datos, incluyendo una imputación provisoria de los datos faltantes. La imputación provisoria es el promedio simple $(\bar{z}_{i.} + \bar{z}_{.j})/2$. El procedimiento es iterativo, aunque suele converger en la primera o segunda iteración. Una vez imputados los datos faltantes, se realizaron sendos ANOVA y *tests* de comparaciones múltiples. Los ANOVA arrojaron diferencias significativas entre pares de cultivos en ambos conjuntos de datos. También arrojaron diferencias entre zonas, pero tales diferencias ya habían sido asumidas al plantear un DBCA. Las pruebas de comparaciones múltiples se realizaron, para cada conjunto de datos, a través de dos estadísticos de prueba, el de Tukey y el de Duncan.

Estimación de los coeficientes de correlación para la región pampeana

La última etapa del estudio consistió en la obtención de coeficientes de correlación generales para toda la región pampeana. Para ello se planteó un criterio de *consenso* entre los resultados hallados en

apartados anteriores. Si llamamos A_i a la matriz clasificatoria que surge de cada comparación combinada con cada conjunto de datos, e.g. Tukey con el conjunto de soja total, podemos minimizar las discrepancias entre clasificaciones mediante el programa lineal (Williams, 2013 pág. 32-34)

$$\min_e \{0' \hat{\rho} + 1' e^+ - 1' e^-\} \text{ s.a. } \begin{bmatrix} A_1 & I \otimes u' & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ A_4 & 0 & \dots & I \otimes u' \end{bmatrix} \begin{bmatrix} \hat{\rho} \\ e_1^- \\ e_1^+ \\ \vdots \\ e_4^- \\ e_4^+ \end{bmatrix} = \begin{bmatrix} \hat{\rho} \\ \Gamma_1 \\ \vdots \\ \Gamma_4 \end{bmatrix}$$

donde $u' = [-1, 1]$ y \otimes se refiere al producto tensorial de Kronecker.

RESULTADOS

La solución del programa lineal anterior, reordenada en forma de matriz de correlaciones, se transcribe en el cuadro 1. Nótese que todos los elementos de la matriz son positivos, tal como suponen los analistas de riesgo agropecuario. Además, se debe tener presente que los coeficientes del cuadro 1 son conservadores, visto que el estimador que surge del programa lineal es un estimador robusto. El ordenamiento de los cultivos en el cuadro no es casual, sino que han sido ordenados por estación de crecimiento y superposición de áreas productivas en la región pampeana.

	Soja 1era.	Soja 2da.	Maíz	Sorgo	Girasol	Trigo
Soja 1era.	1					
Soja 2da.	0,6945	1				
Maíz	0,5201	0,5201	1			
Sorgo	0,5132	0,3964	0,5201	1		
Girasol	0,3134	0,1167	0,1825	0,2620	1	
Trigo	0,1965	0,1167	0,1825	0,1167	0,1825	1

Cuadro 1. Matriz de correlaciones de consenso entre los rendimientos de cinco cultivos extensivos de la región pampeana.

DISCUSIÓN

La simple inspección del cuadro 1 sugiere que las correlaciones entre los cultivos estivales se hallan en torno a 0,50, a excepción de la correlación entre girasol y los demás cultivos estivales, que se halla entre 0,10 y 0,30. La correlación entre el único cultivo invernal (trigo) considerado en este estudio y los cultivos estivales se ubica entre 0,10 y 0,20. Lógicamente, la correlación más alta se da dentro del cultivo de soja sembrada en distinta época, la que ronda 0,70. A partir de estos resultados es posible construir una matriz de correlaciones *estilizada* de fracciones simples. Para ello ajustamos el modelo de regresión

$$\hat{\rho}_{ij} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \epsilon_{ij}$$

donde $x_1, x_2 = 1/2$ si $i, j = 1$, o $\{x_1, x_2\} = \{i-2, j-2\}$ en caso contrario, y ϵ_{ij} es un término de error aleatorio.

Los subíndices i, j corresponden a las filas y columnas del cuadro 1 cuyo ordenamiento es totalmente arbitrario. Se omitió este término por no ser significativo. El cuadro 2 muestra las correlaciones estimadas a partir de las regresiones, redondeados al primer decimal. Los coeficientes del cuadro 2 deben interpretarse en la práctica como cotas inferiores de los coeficientes zonales homólogos, ya que es sabido (Goodwin & Mishra, 2002) que las correlaciones entre cultivos próximos entre sí son mayores que entre cultivos más alejados.

	Soja 1era.	Soja 2da.	Maíz	Sorgo	Girasol	Trigo
Soja 1era.	1,0					
Soja 2da.	0,7	1,0				
Maíz	0,6	0,5	1,0			
Sorgo	0,4	0,4	0,4	1,0		
Girasol	0,3	0,2	0,3	0,3	1,0	
Trigo	0,2	0,1	0,1	0,1	0,2	1,0

Cuadro 2. Matriz de correlaciones de consenso estilizada entre rendimientos de cinco cultivos extensivos de la región pampeana.

Algunos programas informáticos de simulación, como @Risk realizan pruebas de “auto-consistencia” sobre las matrices provistas por el usuario. Una de ellas sugiere que, si la correlación entre el cultivo i y el cultivo j es ρ_{ij} , y la correlación entre el cultivo i y el cultivo k es ρ_{ik} , entonces la correlación entre j y k debe estar acotada en el intervalo general

$$\rho_{jk} \in \rho_{ij}\rho_{ik} \pm \sqrt{(1-\rho_{ij}^2)(1-\rho_{ik}^2)}.$$

El lector puede verificar que la matriz del cuadro 2 satisface esta especificación. La verificación es algo tediosa, pero puede agilizarse notando que si ρ_{jk} pertenece al intervalo

$$-1 + (\max\{\rho_i\}^2 - \min\{\rho_i\}^2) \leq \rho_{jk} \leq 1 - (\max\{\rho_i\}^2 - \min\{\rho_i\}^2)$$

también pertenece al intervalo general. En nuestro caso, inclusive, la cota inferior puede igualarse a cero dado que se espera que todas las correlaciones sean positivas. Lógicamente, aquellas correlaciones que no pertenezcan a este último intervalo deberán chequearse directamente con el intervalo general. En el caso de la matriz del cuadro 2 se verifica que todas las correlaciones, salvo ρ_{21} (soja 1era.-soja 2da.) y ρ_{31} (soja 1era.-maíz), pertenecen al intervalo de cálculo simplificado y que tanto ρ_{21} como ρ_{31} pertenecen al intervalo general. En definitiva, ante la ausencia de información zonal o local, la matriz del cuadro 2 es ampliamente recomendada como primera aproximación.

REFERENCIAS

- Fisher, R. A. (1934). Statistical Method For Research Workers. Fifth Edition. Oliver & Boyd. Edinburgh.
 Goodwin, B. K. & Mishra, A. K., (2002). An Empirical Evaluation Of Yield Performance And Cross-Crop Yield Correlation. 2002 Annual Meeting of the Western Agricultural Economics Association.

Ciudad de Buenos Aires, octubre 2024



July 28-31, Long Beach, California. Disponible en:

<https://ideas.repec.org/p/ags/waealb/36550.html>

Greene, W. (2008). *Econometric Analysis*. 6ta edición. Pearson Prentice Hall, Upper Saddle River.

Gujarati, R. & Porter, D. (2009). *Econometría*. Quinta edición. McGraw-Hill. México.

Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley & Sons Ltd. Baffins Lane, Chichester.

Vose D., (2000). *Risk Analysis. A Quantitative Guide*. Second Edition. John Wiley & Sons Ltd. Baffins Lane, Chichester.

Yates, F. (1933). The Analysis of Replicated Experiments when the Field Results are Incomplete, *Emp. Jour. Experi. Agri.* 129-142.

Williams, H. P. (2013). *Model Building in Mathematical Programming*. 5th Edition. John Wiley & Sons Ltd. West Sussex.

INTERPOLACIÓN ESPACIAL DE MATERIA ORGÁNICA DEL SUELO A ESCALA DE LOTE USANDO QUANTILE REGRESSION FOREST

GARCÍA SELEME F.A.^{1,2}, CÓRDOBA M.A.^{1,2}, PACCIORETTI P.^{1,3,4}, BALZARINI M.^{1,2}

¹ *Cátedra de Estadística y Biometría, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba*

² *Grupo de Estadística Vinculado, Unidad de Fitopatología y Modelización Agrícola. CONICET (UFyMA-INTA-CONICET). Córdoba.*

³ *Comisión Nacional de Actividades Espaciales. Universidad Nacional de Córdoba. Instituto Mario Gulich. Córdoba.*

⁴ *University of Nebraska. Department of Agronomy and Horticulture. Lincoln, Estados Unidos*

fabrizio.garcia.seleme@mi.unc.edu.ar

RESUMEN

La materia orgánica del suelo (MO) es un componente esencial para la fertilidad y productividad agrícola, por lo que estimar su distribución espacial con precisión es esencial para un manejo eficiente y sostenible del suelo. Tradicionalmente, la interpolación espacial se ha realizado mediante métodos geoestadísticos como Kriging, que requieren cumplir supuestos distribucionales y contar con una cantidad suficiente de datos para ajustar semivariogramas confiables. Los algoritmos de aprendizaje automático, como Quantile Regression Forest (QRF), ofrecen una alternativa prometedora al ser más flexibles respecto a estos supuestos y facilitar la integración de datos provenientes de múltiples lotes agrícolas. Sin embargo, QRF no considera la estructura de autocorrelación espacial en el ajuste del modelo. Para abordar esta limitación, han surgido nuevas alternativas como Quantile Regression Forest Spatial Interpolation (QRFI), que incorpora la información espacial utilizando como covariables los valores de la variable respuesta observada en los sitios más cercanos a los puntos de predicción. En este trabajo, se comparó QRFI con Kriging en su capacidad para mapear la distribución intralote de MO. Se utilizaron 163 puntos georreferenciados de MO, muestreados en cinco lotes agrícolas. Los modelos se ajustaron utilizando como variables explicativas la conductividad eléctrica aparente del suelo, elevación y rendimiento del cultivo, procesadas en una grilla común de 10 × 10 m. El error de predicción se estimó mediante validación cruzada k-fold (k=10). Los resultados mostraron que QRFI superó a Kriging en términos de error de predicción (RMSE). QRFI se representa como una herramienta prometedora para el mapeo digital del suelo a escala fina.

Palabras clave: **aprendizaje automático, autocorrelación espacial, mapeo digital del suelo**

INTRODUCCIÓN

El contenido de materia orgánica del suelo ha sido considerado tradicionalmente como uno de los factores fundamentales de la fertilidad de los suelos (Bünemann et al., 2018), estando asociado positivamente con rendimientos más elevados. La predicción de esta variable resulta por ello de interés para la producción agrícola. Con el desarrollo de técnicas de agricultura de precisión, el desarrollo de métodos de predicción más precisos se ha vuelto un aspecto central, permitiendo un manejo eficiente y sostenible del suelo (Kweon & Maxton, 2023).

El método tradicional para realizar estas predicciones es el Kriging en sus distintas modalidades. Este procedimiento requiere del cumplimiento estricto de una serie de supuestos, como la distribución normal de la variable respuesta y la estacionalidad del proceso espacial que genera los datos, como así también una cantidad suficiente de información inicial.

Recientemente se han desarrollado algoritmos de aprendizaje automático especializados en la predicción, como los modelos de bosques aleatorios (RF) (Breiman, 2001). Si bien RF no considera de forma directa la autocorrelación espacial de los datos, experiencias recientes desarrollaron una alternativa que incorpora la estructura espacial de los datos al modelo de predicción, denominada Quantile Regression Forest (QRF) (Córdoba & Balzarini, 2021). Estudios previos probaron la superioridad de QRF sobre KG para predecir valores de rendimiento a partir de datos de monitor de cosechadora (Suárez et al., 2022). Sin embargo, el uso de esta herramienta para la predicción de variables edáficas se encuentra aún poco difundido. Nikou y Tziachris (2022) evaluaron la herramienta para el mapeo de materia orgánica considerando la concentración de componentes minerales como covariables, obteniendo un error cuadrático medio (MSE) superior al conseguido con métodos de Kriging Regresión (RK).

El objetivo de este trabajo fue comparar el método Quantile Random Forest Spatial Interpolation (QRFI) con Kriging en su capacidad para mapear la distribución intralote del contenido de materia orgánica.

MATERIALES Y MÉTODOS

Los lotes utilizados en el presente estudio se ubican en el centro sur de la provincia de Córdoba. Se emplearon cinco lotes. Se cuenta con datos de muestras de suelo georreferenciadas en las cuales se midió contenido de materia orgánica (en %) junto a otras variables que no fueron utilizadas. Además, se contó con información del monitor de cosecha para cada uno de los lotes, sobre cultivo de maíz (*Zea mays* L.), correspondiente a la campaña 2022/23. Por último, de cada lote se posee información de un muestreo realizado con rastra veris para las variables conductividad eléctrica superficial y subsuperficial, al igual que datos de altimetría. Los datos de rendimiento, conductividad y altitud fueron procesados en una grilla de 10x10 metros empleando Kriging Universal ajustado por coordenadas (latitud o longitud) cuando fuera requerido. Una vez obtenidas las mallas de predicción se añadieron a la base de datos de las muestras de suelo los valores predichos para las covariables. Para la predicción del contenido de materia orgánica en cada lote se ajustaron 3 modelos: 2 modelos QRFI, uno global (QRFI_G) que consideró los datos de todos los lotes y uno local (QRFI_L) que consideró únicamente los datos del lote objetivo; y un modelo Kriging Regresión (RK). Los tres

modelos consideraron todas las covariables previamente citadas. Los valores de los datos vecinos fueron incorporados a los modelos QRFI mediante un término de ponderación inversa a la distancia (IDW). Los modelos fueron evaluados mediante validación cruzada k-fold ($k=10$) y comparados entre sí utilizando la raíz del error cuadrático medio (RMSE). Para evaluar la similitud entre mapas se realizaron comparaciones de a pares con prueba T modificada para cada lote ($\alpha = 0.05$). Para los análisis realizados se utilizaron las librerías *autopmap*, *caret*, *gstat*, *ranger*, *SpatialPack* y *sf* de R.

RESULTADOS y DISCUSIÓN

El modelo QRFI_L fue el que menor error registró en todos los lotes utilizados en este estudio (Tabla 1), superando al modelo global aún en los lotes con menor cantidad de muestras. La heterogeneidad del suelo de la zona podría haber generado variaciones entre lotes que hagan poco representativos los datos de diferentes lotes para la predicción. Por otro lado, el modelo RK fue el que mostró mayor RMSE en cuatro de los cinco lotes, presentando valores especialmente altos en los lotes con menor número de muestras, a pesar de ser lotes de menor superficie. La alta variabilidad de las covariables medidas en algunos de estos lotes (ej. Conductividad eléctrica CV = 60% en el lote C frente a variaciones del 20% en otros lotes) sumado a la potencial menor representatividad de dicha variación por menor tamaño muestral pudieron contribuir a estos resultados.

El RMSE de los modelos QRFI_G, QRFI_L y RK en el lote A fue de .242, .208 y .263, respectivamente. En lo que refiere al lote B los RMSE fueron .250, .234 y .235, respectivamente. Para el lote C los RMSE fueron 0.226, 0.203 y 0.312, respectivamente. Para el lote D los RMSE fueron de 0.328, 0.307 y 0.462. En el lote E los RMSE fueron de 0.298, 0.168 y 0.330.

Los resultados son opuestos a lo hallado por Nikou y Tziachris (2022), quienes reportaron que RK daba menores valores de RMSE respecto a un QRF. Numerosos aspectos difieren entre dicha experiencia y la presente, principalmente en cuanto a las covariables utilizadas, el uso del suelo explorado, la densidad de muestreo aplicada y la metodología de validación empleada.

Lote\Modelo	QRFI_G	QRFI_L	RK
A	.242	.208	.263
B	.250	.234	.235
C	.226	.203	.312
D	.328	.307	.462
E	.298	.168	.330

Tabla 1. Error de predicción (RMSE) para tres modelos ajustados en cinco lotes, obtenidos mediante validación cruzada ($k\text{-fold} = 10$).

Respecto a la correlación entre las predicciones de los distintos modelos, la misma fue significativa y altamente positiva para todos los modelos en el lote A ($r > 0.75$ para todas las combinaciones, $p < 0.001$), al igual que para el lote B (ídem lote A). En los lotes C y D no se encontró una correlación significativa entre las predicciones, mientras que para el Lote 36 sí se halló correlación significativa

positiva para las predicciones de RK y ambos QRFI (QRFI_G-RK = 0.60; QRFI_L-RK = 0.53; $p < 0.05$), pero no así para los modelos QRFI entre sí.

Las relaciones entre estas predicciones son apreciables al generar mapas de predicción en la grilla 10 x 10 utilizada para la predicción de las covariables (Figura 1). En el lote A se observa una franja central con valores más bajos de MO, la cual se corresponde con valores de altitud más elevados (sector de loma). En el modelo QRFI_L del lote D se puede apreciar una concentración de valores elevados de MO en torno al sector irregular del perímetro, lugar que coincide con un bajo pronunciado del lote. El conocimiento ampliamente difundido sobre las propiedades edáficas que tienen estas zonas de relieve permite verificar desde lo teórico la validez de las predicciones generadas.

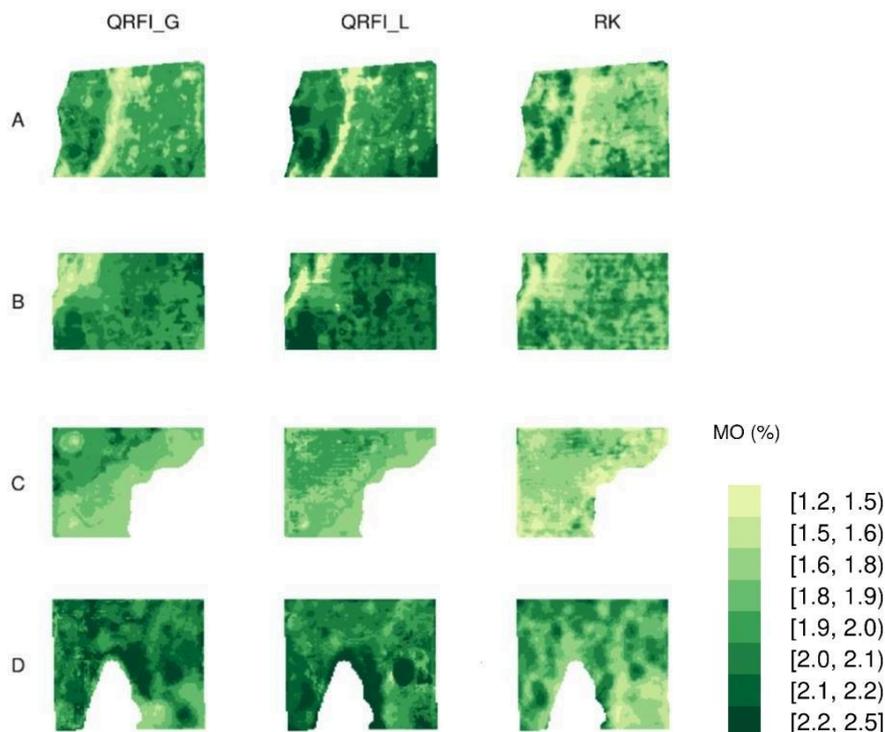


Figura 1. Mapas obtenidos para los modelos ajustados para la predicción del contenido de materia orgánica (MO) para los lotes A, B, C y D.

REFERENCIAS

- BREIMAN, L. (2001). Random forests. *Machine Learning*, 45(1): 5–32.
<https://doi.org/10.1023/A:1010933404324>
- BÜNEMANN, E. K.; BONGIORNO, G.; BAI, Z.; CREAMER, R. E.; DE DEYN, G.; DE GOEDE, R.; BRUSSAARD, L. (2018). Soil quality—A critical review. *Soil Biology and Biochemistry*, 120: 105-125.
- CÓRDOBA, M.; BALZARINI, M. (2021). A random forest-based algorithm for data-intensive spatial interpolation in crop yield mapping. *Computers and Electronics in Agriculture*, 184.
<https://doi.org/10.1016/j.compag.2021.106094>
- KWEON, G., MAXTON, C. (2013). Soil organic matter sensing with an on-the-go optical sensor. *Biosystems engineering*, 115(1): 66-81.



NIKOU, M.; TZIACHRIS, P. (2022). Prediction and Uncertainty Capabilities of Quantile Regression Forests in Estimating Spatial Distribution of Soil Organic Matter. *ISPRS Int. J. Geo-Inf.*, 11, 130. <https://doi.org/10.3390/ijgi11020130>

SUÁREZ, F.; PACCIORETTI, P.; CÓRDOBA, M.; BALZARINI, M. (2022). Comparación de métodos de aprendizaje automático para el mapeo de la variabilidad intralote. En: *2° Congreso Latinoamericano de Agricultura de Precisión – CLAP 2022*. (2ª Ed). Río Cuarto: Gabriel Espósito



USO DE COVARIABLES CLIMÁTICAS Y EDÁFICAS PARA LA DEFINICIÓN DE "AMBIENTE" EN ESTUDIOS DE ESTABILIDAD DEL RENDIMIENTO DE MAÍZ EN ARGENTINA

FEDERICO IZZO¹, CRISTINA CUESTA²

¹*Breeding, Bayer*

²*Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario*

federico.izzo@bayer.com

RESUMEN

La estabilidad en el rendimiento es una de las características más valoradas en genotipos de maíz. Para lograr el mejoramiento genético para estabilidad, se debe cuantificar la interacción genotipo por ambiente (IGA). Para esto se han propuesto diferentes metodologías, pero son pocas en las que la definición de ambiente permite una interpretación conceptual de la IGA. El objetivo de este trabajo fue definir “ambiente” a partir de variables edáficas y climáticas, permitiendo comprender las bases de la IGA en un programa de mejoramiento genético de maíz. Los resultados indican que esta redefinición del factor ambiente, podría permitir obtener una caracterización de los genotipos más consistente para estabilidad en el rendimiento.

Palabras clave: ***interacción genotipo ambiente, mejoramiento genético***

INTRODUCCIÓN

La presencia de IGA es uno de los principales obstáculos para la selección de los mejores genotipos en un programa de mejoramiento genético. Dichas interacciones dan lugar a cambios de ranking para el carácter analizado según el ambiente explorado (interacciones cualitativas). Numerosos autores han propuesto índices para cuantificar la IGA, dando lugar a estudios llamados “análisis de estabilidad”. Esos estudios se agrupan según la metodología estadística utilizada, siendo las más populares las basadas en regresión lineal, análisis multivariados o modelos mixtos.

El carácter de mayor interés para el estudio de la estabilidad en maíz es el rendimiento. En gran parte de los trabajos realizados sobre este carácter, el factor ambiente es determinado como el promedio del rendimiento de todos los genotipos en cada combinación de año y localidad (“índice ambiental”), es decir, en cada ensayo comparativo de rendimiento (ECR) realizado. Un genotipo se considera estable si su respuesta al ambiente es igual a la respuesta promedio de todos los genotipos (Lin et al. 1986) o si la suma de cuadrados del error de su regresión es baja (Lin et al. 1986). Esto tiene consecuencias en el análisis estadístico (falta de cumplimiento de supuestos en análisis de regresión, por ejemplo), y en la capacidad de dar sentido conceptual a la interacción.

Ciudad de Buenos Aires, octubre 2024

Debido a que diferentes genotipos de maíz pueden tener distintos niveles de tolerancia a factores ambientales, un promedio de rendimiento puede ser el resultado de combinaciones de variables edáficas o climáticas muy diferentes. Esto podría significar que la caracterización de un genotipo como “estable” bajo esta definición de ambiente sea poco repetitiva al sumar nueva información. El objetivo de este trabajo es definir “ambiente” de una forma que provea bases conceptuales para el estudio de la IGA. Para esto, se propone sustituir la definición de ambiente como el promedio de los rendimientos de todos los genotipos en cada ECR por una combinación de variables climáticas y edáficas, a través del uso de análisis de *clusters*. De completar este objetivo, se podría utilizar la nueva definición de ambiente para identificar genotipos estables mediante una metodología que permita obtener resultados con interpretación biológica y mayor repetitividad.

METODOLOGÍA

Durante 3 campañas se realizaron ECRs en 36 localidades distribuidas en la región templada Argentina, donde se midieron datos climáticos y edáficos. Los primeros se obtuvieron de estaciones meteorológicas ubicadas en cada ECR. Los segundos se tomaron en puntos al azar en cada ECR. La construcción de los clusters se realizó utilizando las variables edáficas y climáticas. La medida de similaridad elegida fue la distancia Euclidiana. Las variables fueron estandarizadas, de forma que la magnitud de la escala no tenga un efecto sobre el agrupamiento (Anderberg, 1975). Como metodología de agrupamiento, se utilizó el algoritmo de Ward (Ward (1963) y Ward y Hook (1963)), el cual puede ser formalizado como: suponiendo una muestra de n observaciones que ha sido particionada en g grupos, con el grupo i -ésimo conteniendo n_i observaciones con media \bar{x}_i , la suma de desviaciones al cuadrado intra-grupo para el grupo g está dada por:

$$W = \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

Donde x_{ij} es la j -ésima observación en el grupo i -ésimo. La varianza conjunta intra-grupo está dada por el cociente W/n . Entonces, en cada iteración del algoritmo el procedimiento es: (i) Comenzar con n grupos (cada uno con una observación, por lo tanto, $W = 0$) (ii) En cada iteración se reduce el número de grupos en uno, uniendo los dos grupos cuya combinación resulte en el menor aumento de W posible. (iii) Las iteraciones continúan hasta que las uniones de $n-1$ grupos den por resultado un único grupo conteniendo todas las observaciones.

Una de las principales características de los algoritmos de agrupamiento jerárquico es que no requiere la pre-definición del número de *clusters* a formar. Para definirlo, se utiliza un dendograma, mediante el cual Anderberg (1975) recomienda tomar como punto de corte un valor de similaridad que forme grupos conceptualmente significativos. La evaluación de los *clusters* se realiza a partir de variables que no fueron utilizadas para generarlos, como rendimiento, ubicación geográfica, fecha de siembra y campaña, y por variables utilizadas para la construcción de los mismos.

Finalmente, se comparan los resultados de calcular la estabilidad de 50 genotipos de maíz incluidos en los ECRs mediante un modelo AMMI, tanto con el factor ambiente definido como un promedio de rendimientos como definido mediante los clusters. Los resultados se representan en gge-biplots.

Los análisis se realizan en el software R (R Core Team, 2022). Se utiliza la función “hclust” del paquete

Ciudad de Buenos Aires, octubre 2024

“Stats” (R Core Team, 2022) y “gg_biplot” del paquete “stability” (Yaseen et al. 2018).

RESULTADOS Y DISCUSIÓN

La figura 1 presenta el dendrograma resultante de aplicar el algoritmo de Ward. Puede observarse por los valores de similitud que inicialmente hay dos grupos muy diferenciados de ambientes.

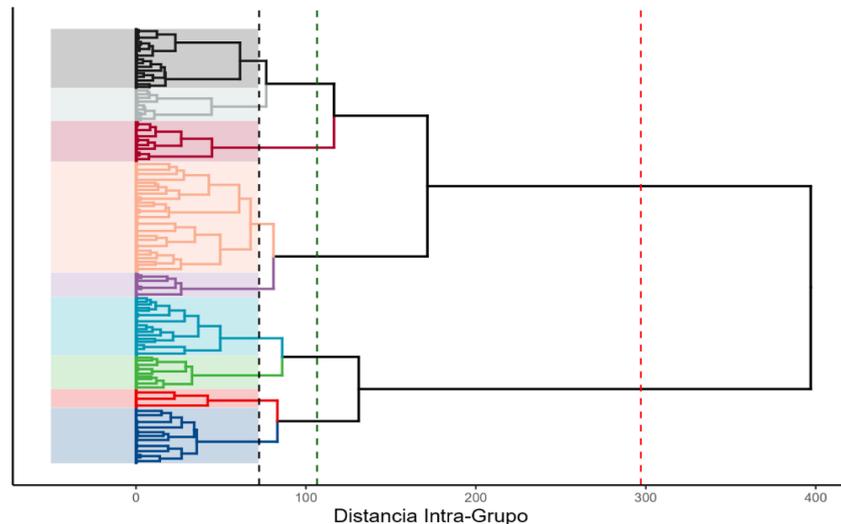


Figura 1: Dendrograma las líneas marcan los valores de corte. Los rectángulos en colores indican los clusters

Esta diferenciación de ambientes puede ser explicada completamente por la fecha de siembra (división de clusters a partir de la primera línea observando el gráfico de derecha a izquierda). En la siguiente subdivisión, quedaron diferenciados ambientes correspondientes en principio a fecha de siembra y en segundo lugar a las diferentes campañas agrícolas, esta subdivisión corresponde a la distancia indicada por la segunda línea entrecortada. Al agrupar los ambientes en nueve *clusters*, quedaron formados grupos por fecha de siembra, campaña, y sub-regiones geográficas. Divisiones subsiguientes dieron lugar a *clusters* con pocas observaciones, por lo que se decidió conservar este número para representar los ambientes.

La figura 2 presenta la distribución del rendimiento en cada uno de los clusters formados. Pueden observarse clusters relacionados a rendimientos medios/altos con baja variabilidad (5, 6 y 7) (clusters con altas precipitaciones o fechas de siembra tardías), rendimientos medios/altos con mayor variabilidad (4, 3 y 2) (precipitaciones normales y fechas de siembratempranas), rendimientos muy variables (1), y rendimientos bajos (8 y 9) (*clusters* caracterizados por muy bajas precipitaciones).

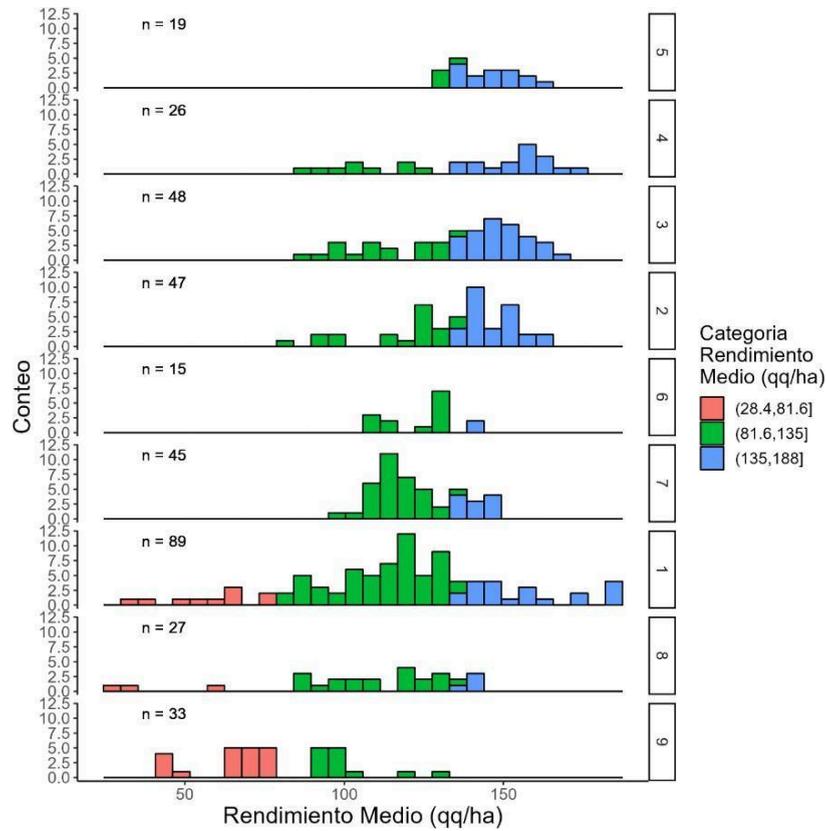


Figura 2: distribución del rendimiento de los ECRs (qq/ha) por cluster

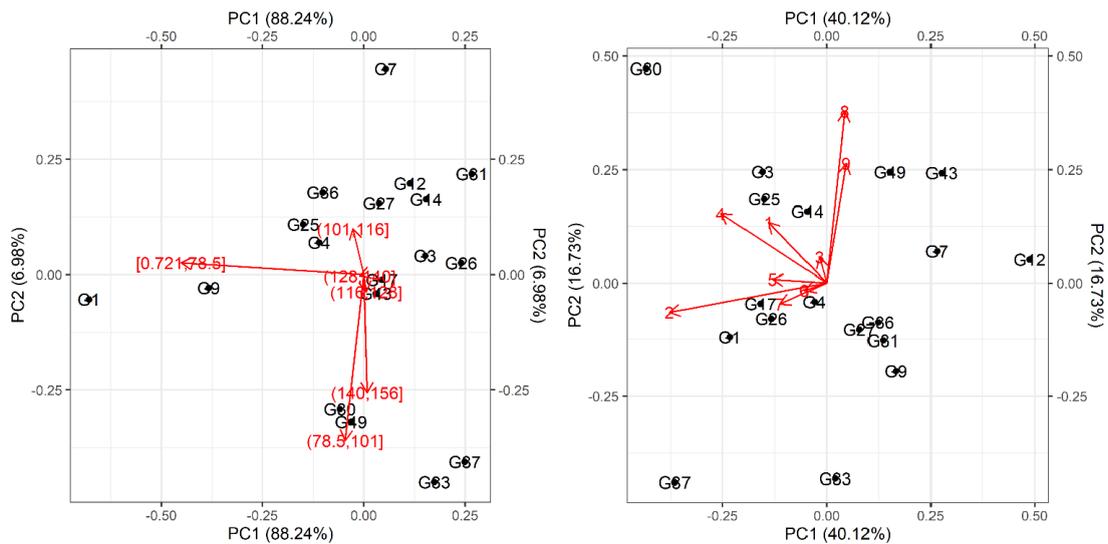


Figura 3: biplots de genotipos y ambientes: 3.a (izq.) ambiente como promedio de rendimiento y 3.b (der.) ambiente definido por los clusters

Para contrastar los resultados de modificar la definición de ambiente de índice ambiental (figura 3.a) a clusters (figura 3.b), se estima la estabilidad de genotipos mediante el ajuste de GGE Biplots. La figura 3.a muestra genotipos adaptados a ambientes de rendimiento particulares, por ejemplo, el genotipo 49 muestra adaptación a ambientes de 78.5 a 101 quintales por hectárea, pero este resultado podría no ser necesariamente repetitivo (el mismo ambiente de rendimiento puede deberse a combinaciones de factores abióticos y bióticos diferentes). También se puede observar que el genotipo 4 muestra adaptación a ambientes de 101 qq/ha a 116 qq/ha.

La figura 3.b muestra que el genotipo 49 muestra adaptación a ambientes de pocas precipitaciones y altas temperaturas (representados por los clusters 8 y 9), este resultado sí es consecuencia de factores que definen el rendimiento, por lo que, de sumar nueva información, sería esperable que sea repetitivo. Al mismo tiempo, puede observarse que el genotipo 4 es el de mayor estabilidad, sin mostrar adaptación a un ambiente específico, lo cual contrasta con los resultados observados para este mismo genotipo en análisis con el factor ambiente definido por rendimientos.

Los resultados demuestran que la definición de ambiente a partir de *clusters* de variables climáticas y edáficas pueden aportar interpretación al estudio de la IGA. Caracterizar genotipos de esta forma puede mejorar la repetitividad de los resultados, mejorando la capacidad de los programas de mejoramiento de seleccionar para estabilidad en el rendimiento a través de los años.

REFERENCIAS

- ANDERBERG, M.R. (1973): *Cluster Analysis for Applications*. Academic Press, New York.
- FAO. 2023. *Datos sobre alimentación y agricultura*. [[https:// www.fao.org](https://www.fao.org)] Licencia: CC-BY-4.0". Roma, Organización de las Naciones Unidas para la Agricultura y la Alimentación.
- LIN, C. S., BINNS, M. R., & LEFKOVITCH, L. P. (1986): *Stability analysis: where do we stand?* 1. *Crop science*, 26(5)
- MUHAMMAD YASEEN, KENT M. ESKRIDGE AND GHULAM MURTAZA (2018): *Stability Analysis of Genotype by Environment Interaction (GEI)*. R package version 0.5.0. <https://CRAN.R-project.org/package=stability>
- R CORE TEAM (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- WARD JR, J. H., & HOOK, M. E. (1963): *Application of an hierarchical grouping procedure to a problem of grouping profiles*. *Educational and psychological measurement*, 23(1), 69-81

CLASIFICACIÓN DE LA CALIDAD DEL HÁBITAT EN CURSOS DE AGUA DE LA PROVINCIA DEL NEUQUÉN A PARTIR DEL ÍNDICE IVH

LAFFITTE LORENA¹⁻⁴, ABELLI BONARDI MARISOL², FERNÁNDEZ JOSÉ CARLOS¹, GARCÍA TOURN ANDREA², LAURENZANO BETINA², LILLO MANUEL¹, MUÑIZ SAAVEDRA JULIETA⁴, RIVERA DANIEL¹, TORRES MARIANA¹, TRONCOSO LEANDRO², UGOLINI FLORENCIA² Y LAVALLE ANDREA³

¹ Secretaría de Ambiente de la Provincia del Neuquén

² Subsecretaría de Recursos Hídricos de la Provincia del Neuquén

³ Facultad de Economía, Departamento de Estadística, Universidad Nacional del Comahue, Neuquén

⁴ CRUSMA, Universidad Nacional del Comahue, Neuquén.

laffittesmandes@gmail.com

RESUMEN

El trabajo tuvo como objetivo establecer categorías de calidad del hábitat, a partir de la aplicación del Índice de Valoración del Hábitat (IVH) en cursos de agua de la provincia del Neuquén. Se utilizó información de 12 años de muestreos realizados en 93 sitios distribuidos en las cuencas de los ríos Limay, Neuquén, Hua Hum y Colorado. Se incluyeron variables fisicoquímicas como conductividad eléctrica, oxígeno disuelto, nitrógeno total, fósforo total y alcalinidad, además de las 10 variables relacionadas con la caracterización del hábitat y el índice biótico BMPS. Para determinar los rangos que definen las categorías de calidad del IVH, se aplicó un Análisis de Componentes Principales (ACP) utilizando 16 variables descriptivas. La primera componente principal explicó el 39,08% de la variabilidad total y mostró una correlación de 0,96 con el IVH, lo que permitió clasificar los sitios en cuatro grupos. Se establecieron los siguientes puntos de corte para las categorías de calidad del hábitat: 0 a 70 (pobre), 71 a 120 (marginal), 121 a 139 (subóptimo) y 140 a 200 (óptimo). La clasificación fue validada mediante un Análisis Discriminante, obteniéndose una tasa de error del 4%, lo que respalda la robustez de los resultados. La clasificación de la calidad del hábitat en diferentes categorías proporciona una herramienta que permite identificar su condición general y realizar un seguimiento más claro de la salud de los ambientes acuáticos a lo largo del tiempo.

Palabras clave: ***análisis multivariado, valoración del hábitat, índice, macroinvertebrados***

INTRODUCCIÓN

En el año 2011, se comenzó a implementar la evaluación del hábitat en la provincia del Neuquén con protocolos adaptados de los Protocolos de Bioevaluación Rápida desarrollados por la USEPA (Barbour et al., 1999; Flotemersch et al., 2006), tras la primera detección del alga invasora *Didymosphenia geminata*, al establecerse el Programa de Vigilancia y Seguimiento con el objetivo de monitorear su dispersión, incluyendo además el muestreo físico-químico de las aguas, la medición de parámetros de campo, el muestreo de fitoplancton y perifiton (multihábitat). El monitoreo permanente se lleva adelante por dependencias técnicas de Recursos Hídricos y Ambiente de la provincia. A su vez, en 2013, se incorporó al monitoreo de más de 70 sitios, el muestreo de macroinvertebrados acuáticos para utilizarlos como bioindicadores de la calidad del agua, utilizando los mismos protocolos de la USEPA.

El objetivo del trabajo fue establecer categorías de calidad del hábitat, a partir de la aplicación del Índice de Valoración del Hábitat (IVH) en cursos de agua de la provincia del Neuquén, validando los resultados obtenidos mediante análisis estadísticos, y contribuyendo al uso de este índice para la evaluación específica del hábitat. Dicho índice proporciona una medida cuantificable y estandarizada que facilita la comparación entre diferentes cursos de agua y permite detectar cambios en la calidad del hábitat a lo largo del tiempo.

METODOLOGÍA

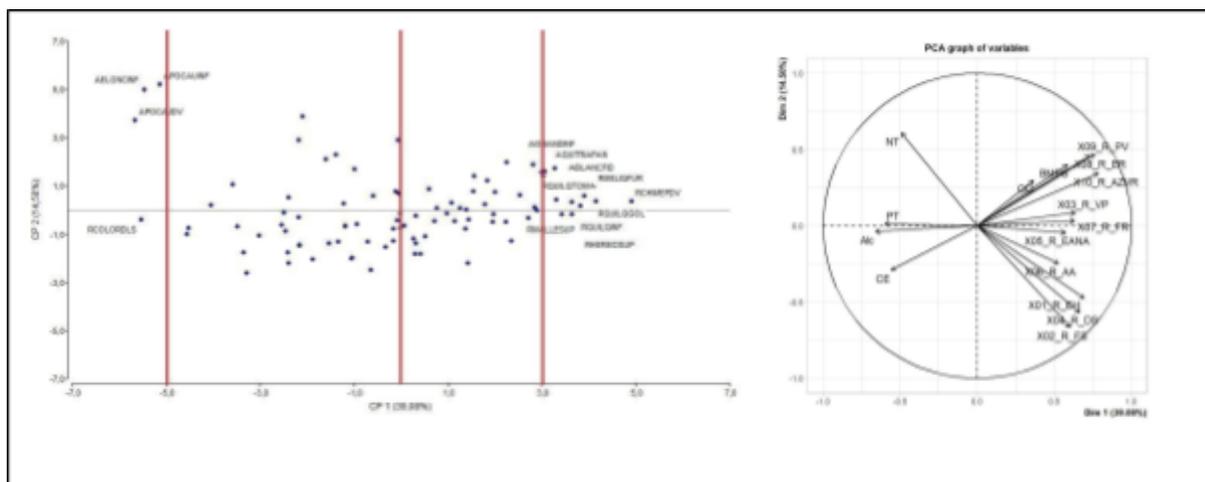
Los resultados surgen del monitoreo de 93 sitios muestreados en las cuencas de los ríos Limay, Neuquén, Hua Hum y Colorado, durante un periodo de muestreo de 12 años siguiendo la metodología adaptada de los Protocolos de Bioevaluación Rápida (Barbour et al., 1996, 1999; Flotemersch et al., 2006). Asimismo pueden encontrarse protocolos para Argentina en Dominguez et al. (2020) y Giorgi et al. (2022). En este estudio se incluyeron 16 variables descriptivas: 1) Variables físico-químicas in situ [Oxígeno Disuelto - OD, Conductividad Eléctrica – CE; equipo multiparámetro HACH], 2) Análisis químico del agua [NT, PT, Alcalinidad; muestras de agua; Laboratorio GECARA CRUB], 3) Análisis biológico [muestras de macroinvertebrados acuáticos - BMPS; Laboratorio CIRA-UNCO], 4) Caracterización del Hábitat [10 variables incluidas en las planillas IVH]. La caracterización del hábitat con el índice IVH consiste en calificar 10 parámetros: 1) estabilidad del hábitat y desarrollo de la biota, 2) fijación del sustrato, 3) velocidad y profundidad, 4) depositación de sedimentos, 5) estado actual del nivel del agua, 6) alteraciones antrópicas del cauce, 7) frecuencia de rápidos, 8) estabilidad de las riberas, 9) protección vegetal, 10) alteración de la zona con vegetación riparia; en una escala numérica desde 0 a 20. Cada variable se evalúa de forma independiente y la suma total de puntos indica una de las cuatro categorías de calidad del hábitat: óptimo, subóptimo, marginal o pobre. Para determinar las categorías de calidad del IVH (pobre, marginal, subóptimo, óptimo), se aplicó un Análisis de Componentes Principales (ACP) utilizando 16 variables descriptivas que influyen en la calidad del hábitat. Se utilizó el conjunto de datos para identificar las funciones discriminantes que mejor separan las cuatro categorías de calidad del hábitat. Las funciones discriminantes son combinaciones lineales de las variables originales (físico-químicas y del hábitat) que maximizan la separación entre los grupos (categorías). A partir de estas funciones, se calculó la probabilidad de que cada observación (sitio muestreado) perteneciera

Ciudad de Buenos Aires, octubre 2024

a una de las categorías de calidad del hábitat. Luego, se compararon las asignaciones originales de los sitios con las asignaciones realizadas por el Análisis Discriminante, lo que permitió medir la precisión del modelo. Los análisis estadísticos fueron realizados con R.

RESULTADOS Y DISCUSIÓN

La primera componente principal explicó el 39,08% de la variabilidad total y mostró una correlación de 0,9631 con el IVH, lo que permitió clasificar los sitios en cuatro grupos: grupos con las observaciones más extremas, tanto positivas (mejor calidad) como negativas (peor calidad), grupos con valores intermedios, positivos por un lado y negativos por otro.



Se encontró una correlación significativa (0,96) entre la primera componente principal y el IVH, y cómo esto respalda la clasificación de los sitios en cuatro grupos de calidad.

Para la clasificación de la calidad del hábitat se establecieron los puntos de corte estableciéndose las 4 categorías:

Categoría	Rango	CARACTERÍSTICAS
ÓPTIMO 	140-200	Indica condiciones de hábitat excelentes, cercanas a las naturales. El ecosistema muestra mínimas alteraciones y se encuentra en su estado más saludable, con capacidad para soportar una biodiversidad rica y equilibrada.
SUBÓPTIMO 	121-139	Representa un hábitat en buen estado, aunque con algunas alteraciones leves o moderadas. Todavía es capaz de soportar comunidades biológicas diversas, pero se observan ciertos impactos en la estructura del hábitat.
MARGINAL 	71-120	Muestra signos evidentes de degradación del hábitat. Las condiciones del agua y la estructura del entorno están afectadas, lo que reduce la capacidad de soporte de la biota nativa y limita la salud general del ecosistema.

<p>POBRE</p> 	<p>0-70</p>	<p>Refleja un hábitat gravemente degradado, con alteraciones severas. Las condiciones del agua y el entorno son extremadamente adversas para la biota, lo que implica una pérdida significativa de biodiversidad y funciones ecológicas.</p>
--	-------------	--

Para validar la clasificación se realizó un Análisis Discriminante, con una baja tasa de error del 4%, lo que indica que solo el 4% de las observaciones se clasificaron incorrectamente en categorías de calidad distintas a las asignadas originalmente. Esto significa que el modelo estadístico basado en el IVH y las variables utilizadas tiene una alta precisión para discriminar entre los diferentes niveles de calidad del hábitat. La baja tasa de error sugiere que las categorías de calidad del hábitat establecidas son robustas y que las variables incluidas en el modelo son adecuadas para describir y diferenciar correctamente las condiciones de los sitios muestreados.

El propósito final de este trabajo es promover el uso del índice como indicador clave para la evaluación de la calidad de los hábitats acuáticos. Al proporcionar un enfoque cuantificable y estandarizado, el índice no solo facilita la identificación de áreas con alta o baja calidad ambiental, sino que también permite a los gestores y tomadores de decisiones monitorear los cambios en los ecosistemas a lo largo del tiempo, y brindar herramientas para la adecuada gestión de los mismos. Su aplicación puede apoyar la implementación de políticas de conservación más efectivas y guiar las acciones de restauración en áreas degradadas. Además, el uso de este índice puede integrarse en programas de monitoreo a largo plazo, fortaleciendo la capacidad de respuesta ante amenazas ambientales y promoviendo la sostenibilidad de los ecosistemas acuáticos. A futuro es importante incrementar los sitios de muestreo en áreas no abordadas, para evaluar posibles diferencias por regiones fitogeográficas.

REFERENCIAS

- BARBOUR, M.T., GERRITSEN, J., SNYDER, B.D., Y STRIBLING, J.B. (1999) Rapid Bioassessment Protocols for Use in Streams and Wadeable Rivers: Periphyton, Benthic Macroinvertebrates and Fish, Second Edition. EPA 841-B-99-002. U.S. Environmental Protection Agency; Office of Water; Washington, DC.
- DOMÍNGUEZ, E. A. GIORGI Y N. GÓMEZ (COMPS.) (2020). La bioindicación en el monitoreo y evaluación de los sistemas fluviales de la Argentina: Bases para el análisis de la integridad ecológica. Editorial Eudeba.
- GIORGI A., E. DOMÍNGUEZ Y N. GÓMEZ (COMPS.) (2022). Técnicas de monitoreo para ecosistemas fluviales de la Argentina. REM.AQUA (Red de Evaluación y Monitoreo de Ecosistemas Acuáticos), Conicet.
- FLOTEMERSCH, J.E., J. B. STRIBLING, Y M.J. PAUL. (2006). "Conceptos y Enfoques Básicos para la Evaluación Biológica de Arroyos y Ríos no vadeables". EPA 600-R-06-127. US Environmental Protection Agency, Cincinnati, Ohio.
- UGCA. (2012) Protocolo para la detección y seguimiento de *Didymosphenia geminata* (Lyngbye) Schmidt en el área andina patagónica. AIC, SSAyDS, DGBA, DPRH, DPA



RIESGO RELATIVO DE INFECCIONES RESPIRATORIAS Y VIRUS SINCICIAL EN EL AMBA: UNA MIRADA DESDE LA ESTADÍSTICA ESPACIAL PARA ANALIZAR SU DINÁMICA PRE Y POSTPANDEMIA

DANIELA PARADA¹, SILVINA DEL DUCA^{1,2}

¹*Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires*

²*Departamento de Matemática, Facultad de Ciencias Económicas, Universidad de Buenos Aires*

daniela.parada@ic.fcen.uba.ar

RESUMEN

Las Infecciones Respiratorias Agudas Bajas (IRAB), que incluyen la neumonía y la bronquiolitis, son la principal causa de ingresos hospitalarios y de ingresos a la Unidad de Cuidados Intensivos Pediátricos (UCIP), a nivel mundial. Además, estas infecciones son las que más contribuyen al número de muertes de niños menores de 5 años. Varios patógenos respiratorios pueden causar las IRAB, siendo el Virus Sincicial Respiratorio (VSR) el agente etiológico predominante en la población de niños menores de cinco años. La carga de enfermedad por VSR impacta de manera sustancial en los servicios de salud. En este trabajo se presentan los resultados de un análisis exploratorio espacial de los datos de atención pediátrica de IRAB en cuatro hospitales públicos del Área Metropolitana de Buenos Aires (AMBA) entre los años 2018 y 2023. Tales datos fueron provistos por el equipo de investigadores del Centro INFANT de Medicina Traslacional (CIMeT) en el marco de un estudio sobre la dinámica de la infección por VSR en el contexto de COVID-19 en el AMBA, y de un estudio acerca de la correlación espacial de las IRAB severas y moderadas y el Índice de Vulnerabilidad Sanitaria (IVS). En particular, de nuestro trabajo se dependen la estimación de la intensidad de casos, tanto de IRAB como de VSR, así como la estimación espacial del riesgo relativo de la enfermedad por VSR severa para el período pre y postpandemia. Para todos los análisis se utilizó *R* (versión 4.2.3). Los resultados del trabajo aportan al actual estado del arte y pueden ser útiles para detectar oportunidades de intervención, tanto para los formuladores de políticas como para los profesionales de salud.

INTRODUCCIÓN

El Virus Sincicial Respiratorio (VSR) es la principal causa de bronquiolitis y neumonía viral en menores de 12 meses en todo el mundo y la causa más frecuente de hospitalizaciones y admisiones en Unidades de Terapia Intensiva Pediátricas (UTIP). Es también una importante causa de mortalidad infantil, particularmente en el mundo en desarrollo. Más del 90% de los episodios de mortalidad ocurren en países en vías de desarrollo y se estima que más del 50% de estas muertes ocurren a nivel comunitario o fuera de un establecimiento de salud. Como resultado, la carga de enfermedad por VSR impacta de manera sustancial en los servicios de salud. Además, cerca del 45% de los ingresos hospitalarios y las muertes intrahospitalarias por Infecciones Respiratorias Agudas Bajas (IRAB) por VSR ocurren en niños menores de 6 meses de edad.

La nueva enfermedad por coronavirus (COVID-19) detuvo al mundo desde marzo de 2020. Dada la falta de medicamentos o vacunas eficaces contra este virus, se tomaron en los distintos países, entre ellos Argentina, medidas sin precedentes para evitar la propagación y reducir la circulación poblacional del SARS-CoV-2: cuarentena, uso obligatorio de barbijos, distanciamiento físico, concientización masiva acerca de la estricta higiene de manos y superficies y cierre de escuelas y actividades grupales. Estas medidas no sólo tuvieron efecto sobre la propagación del SARS-CoV-2, sino que también demostraron tener influencia sobre la transmisión de otros virus respiratorios, entre ellos el VSR. Durante el año 2020 no se observaron brotes de infección severa por VSR en Argentina ni en la mayoría de los países del mundo. Dicha situación se extendió a la primera mitad del 2021, donde a pesar de brotes aislados de circulación de VSR este no alcanzó las tasas habituales de hospitalización y mortalidad observada hasta el 2019. Esta situación sin precedentes en la historia de la enfermedad por VSR podría generar patrones atípicos de contagios y rebrote, así como cambios en la gravedad de la enfermedad debido a la mayor población susceptible. En consecuencia, el grupo más vulnerable de infección severa (menores de 12 meses) podría ser sometido a una mayor carga de VSR luego del fin de las medidas restrictivas impuestas en nuestro país.

En el marco de un estudio sobre cambios en la dinámica de la infección por VSR pre y post COVID-19 en el Área Metropolitana de Buenos Aires (AMBA) y de otro estudio acerca de la correlación espacial de las IRAB severas y moderadas y el Índice de Vulnerabilidad Sanitaria (IVS), el equipo de investigadores del Centro INFANT de Medicina Traslacional (CIMeT, UNSAM) puso a disposición datos de 9290 casos de atención pediátrica de IRAB en cuatro hospitales públicos del AMBA, entre los años 2018 y 2023.

El objetivo del trabajo es investigar sobre posibles cambios en la dinámica espacial de la infección por VSR entre niños menores de cinco años con IRAB en el AMBA a partir de la pandemia de COVID-19, así como cambios en la gravedad de la enfermedad desde un enfoque espacial, y posibles correlaciones con variables socioeconómicas.

METODOLOGÍA Y ANÁLISIS EXPLORATORIO

Para el análisis, se tomaron 9290 casos de infecciones pediátricas de las vías respiratorias inferiores: los datos fueron hospitalarios, multicéntricos y recopilados de forma prospectiva en 2023. Se obtuvieron de cuatro distritos geoespaciales de la Red Infant-ReMBA, como puede verse en el panel derecho de la Figura 1.

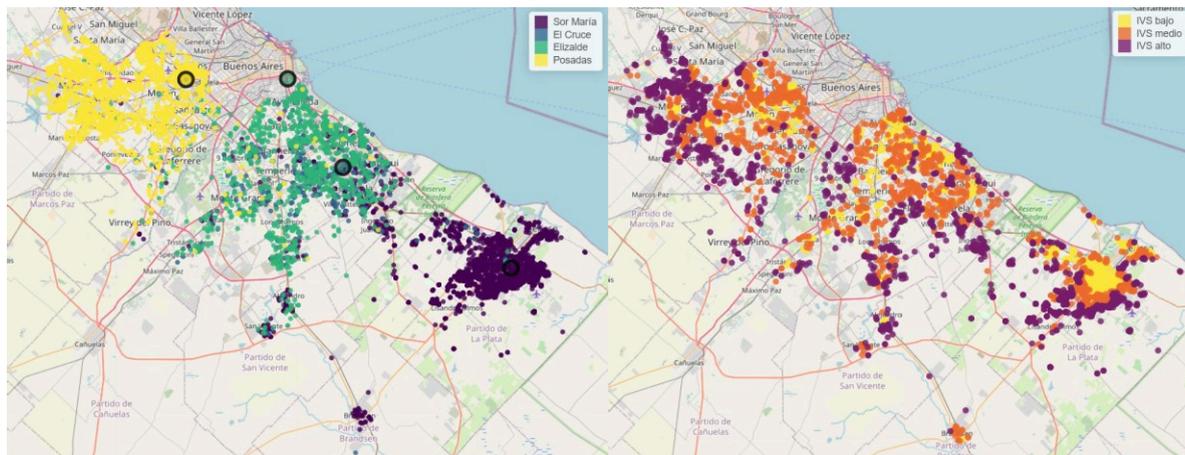


Figura 1. Panel izquierdo. Domicilios georreferenciados (puntos de color) que corresponden a 9290 casos de niños menores de cinco años que ingresaron en uno de los cuatro hospitales de la Red Infant-ReMBA (disco negro) del estudio, entre los años 2018 y 2023, con diagnóstico clínico de IRAB. Los colores amarillo, verde, azul y violeta corresponden a los hospitales Posadas, Elizalde, El Cruce y Ludovica, respectivamente. Panel derecho. Domicilios georreferenciados coloreados según el nivel del Índice de Vulnerabilidad Social (IVS) correspondiente al radio censal del domicilio y estratificado en bajo (amarillo), medio (naranja) y alto (violeta).

1. Ciudad de La Plata, en el extremo sur del AMBA: **Hospital de Niños Sor María Ludovica** (Ludovica);
2. Región Sur del AMBA: **Hospital "El Cruce" de Florencio Varela** (El Cruce);
3. Ciudad de Buenos Aires: **Hospital de Niños Pedro Elizalde** (Elizalde);
4. Región Oeste del AMBA: **Hospital Nacional Dr. Alejandro Posadas** de El Palomar (Posadas).

Se tomó como criterio de inclusión a todos los niños menores de cinco años que ingresaron en uno de los hospitales de la Red Infant-ReMBA, entre los años 2018 y 2023, con diagnóstico clínico de IRAB. Se recopiló la dirección completa, las características basales y los resultados de cada paciente. El de geocodificación con ArcGIS Pro se aplicó a la mayoría de registros, y se complementó con un proceso de geocodificación manual utilizando coordenadas de latitud y longitud extraídas de Google Maps, pudiendo garantizarse la precisión de los datos de ubicación de los pacientes, especialmente para los que residen en regiones socioeconómicamente desfavorecidas y marginadas, donde las herramientas de mapeo digital como Google Earth pueden tener dificultades para proporcionar datos precisos.

El IVS, obtenido a través de los datos disponibles en *poblaciones.org* para cada domicilio y radio censal, es el índice de vulnerabilidad sanitaria y es un indicador que refleja cuán expuesta está una población a riesgos sanitarios y cuán capaces son de enfrentarlos. Se estratificó en bajo, medio y alto, de modo que una población con índice bajo está mejor preparada para atender un riesgo sanitario que una con índice alto.

Se consideró como casos positivos de VSR a todos aquellos que, a lo largo de su estadía hospitalaria, hubiera resultados positivos para alguno de las pruebas de VSR aplicadas (VSR+). La categorización de los casos de IRAB y de VSR+ se definió por la gravedad de su condición durante el ingreso hospitalario. Los casos que requirieron ingreso en la UCIP se clasificaron como graves, mientras que los ingresados en la sala de pediatría se consideran leves o moderados.

El análisis comparativo entre hospitales presentado en la Figura 2 subraya la especialización de cada Ciudad de Buenos Aires, octubre 2024

institución. Los hospitales pediátricos Ludovica y Elizalde muestran una preponderancia de casos leves, mientras que El Cruce, hospital de alta complejidad, concentra la mayor cantidad de casos graves. La segmentación de los pacientes por nivel de gravedad entre estos hospitales no es trivial, pues responde a un patrón espacial y social que no solo está vinculado a la capacidad hospitalaria, sino también a la derivación de pacientes desde otras áreas y hospitales con menor capacidad de respuesta ante casos graves.

La pandemia de COVID-19 introdujo cambios sin precedentes en la dinámica estacional del VSR, tal como se ilustra en las Figuras 3 y 4. Si bien la ausencia casi total de circulación de VSR en 2020 es explicable por las medidas de confinamiento y distanciamiento social, los patrones observados en 2021 y 2022 ofrecen un terreno fértil para el debate. Por un lado, el retraso en el inicio de la temporada de 2021 refleja claramente los efectos persistentes de las restricciones de movilidad y el cierre parcial de las instituciones educativas. Por otro lado, la extensión inusual de la temporada de VSR en 2022 plantea interrogantes sobre la dinámica de la inmunidad poblacional: ¿la falta de exposición al VSR en 2020 y parte de 2021, generó un aumento en la susceptibilidad de los niños, lo que resultó en una mayor transmisión en 2022? No sólo mayor transmisión sino, posiblemente, de mayor gravedad. Este tipo de patrones complejos podría beneficiarse de modelos epidemiológicos más específicos que incorporen efectos de inmunidad de rebaño, posibles cambios en la virulencia viral postpandemia y el impacto de la relajación de las medidas sanitarias.

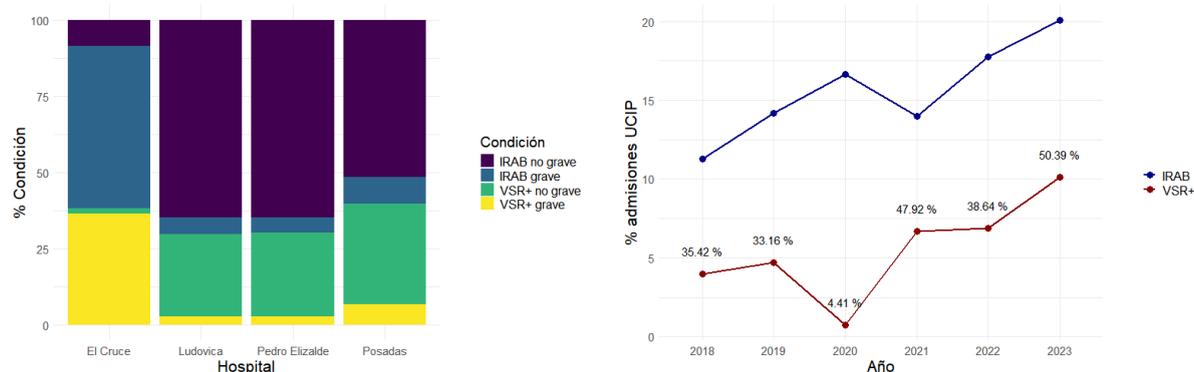


Figura 2. Panel izquierdo. Condiciones de las IRAB según cada hospital. Los hospitales Ludovica y Elizalde, ambos hospitales pediátricos, presentan una distribución de casos similar, con poca presencia de casos graves. El hospital Posadas presentan una tendencia similar, con mayor proporción de casos graves (tanto de IRAB como de VSR). El hospital El Cruce, un hospital de alta complejidad y típicamente de derivación, es el que mayor cantidad de casos de gravedad concentra. Panel derecho. Evolución de la gravedad a lo largo de las temporadas del estudio, tanto en IRAB como en casos de VSR. El porcentaje de admisión a UCIP (severidad) muestra ligeros aumentos a lo largo de las temporadas. Los porcentajes de la curva roja representan la proporción de casos de VSR dentro de las admisiones a UCIP. Se puede notar una significativa la disminución del porcentaje de infecciones respiratorias por VSR durante el año 2020 en el cual da inicio la pandemia y la cuarentena obligatoria.

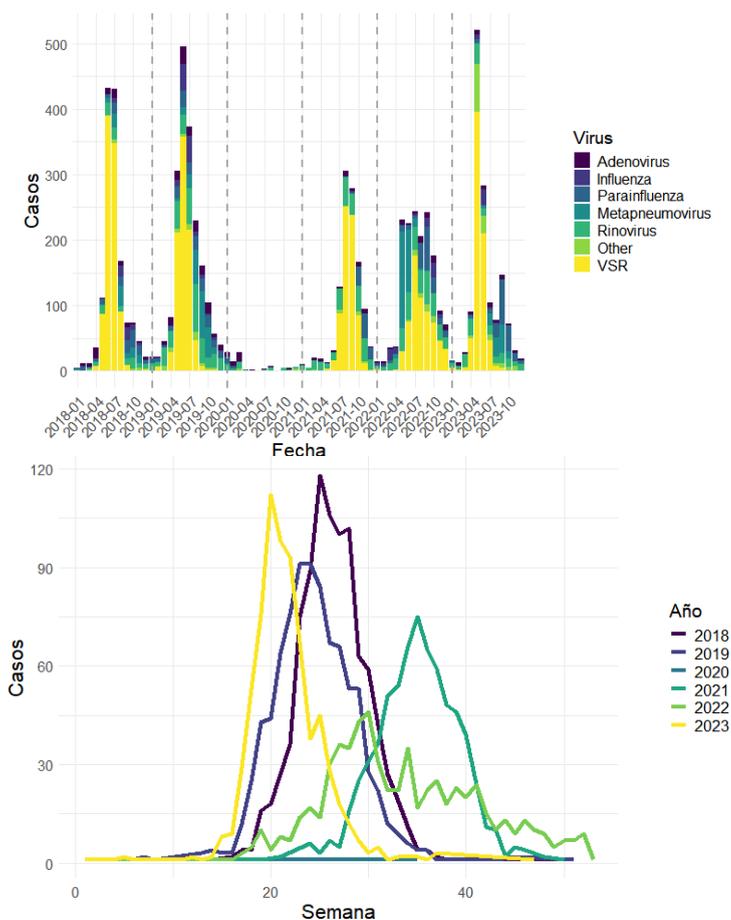


Figura 3. Panel superior. Diagrama de barras apiladas de los casos de virus que resultaron positivos a lo largo de las temporadas. En todos los casos se observa que los casos de VSR (barras amarillas) representan la mayor cantidad de casos a lo largo de todas las temporadas (excepto 2020 e inicios del 2021, con casi nula cantidad de hospitalizaciones por IRAB debido a las restricciones de circulación y confinamiento). Panel inferior. Casos de VSR por temporada. En las curvas se observa el efecto postpandemia en la temporada de sincial: el año 2021 presentó una temporada tardía, con un corrimiento de las semanas epidemiológicas clásicas de sincial (semanas 18 a 32), mientras que el año 2022 presentó una temporada muy larga, que dio lugar a una temporada mucho más anticipada para 2023.

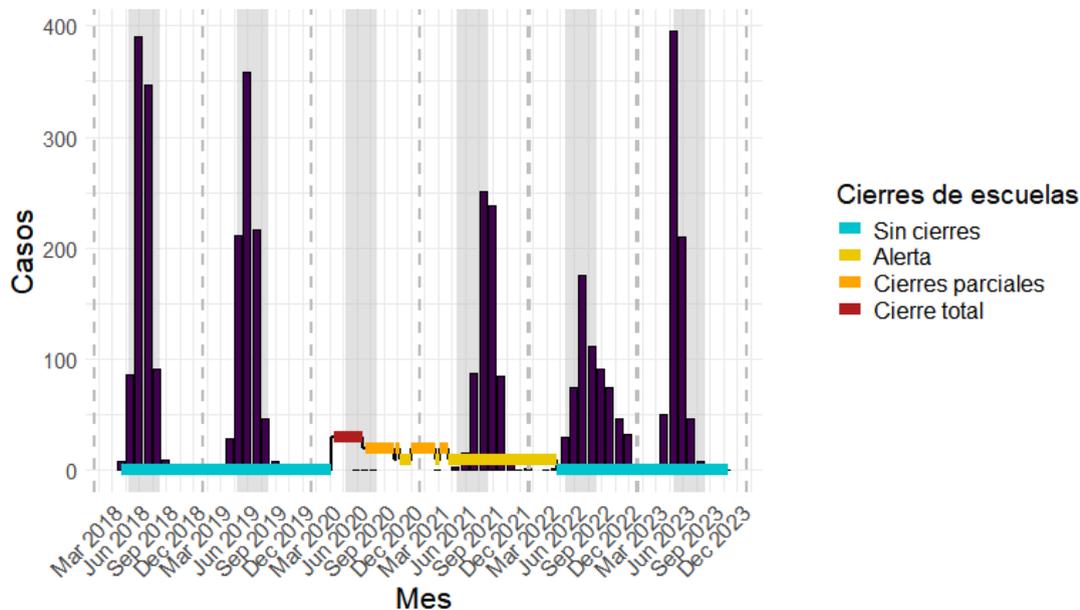


Figura 4. Casos de VSR por mes. En gris se indican las semanas epidemiológicas típicas de sincicial (semanas 18 a 32 del año), lo que es consistente con lo observado para las temporadas de 2018 y 2019. La temporada de sincicial en 2020 fue inexistente, lo que coincide con las diferentes medidas relativas a cierres de escuelas y restricciones de circulación. La temporada 2021 inicia más tarde que lo usual, consistente también con los cierres parciales y alertas de circulación de la época. La temporada 2022 inicia de forma temprana y presenta una duración de casi todo el año. La temporada 2023 inicia ligeramente antes, con una intensidad bastante aguda, similar a las temporadas previas a la pandemia.

Los flujos espaciales de pacientes, ilustrados en la Figura 5, ofrecen un panorama interesante sobre la relación entre geografía, movilidad y acceso a los servicios de salud. El análisis de los vectores de flujo revela que los hospitales no presentan una distribución simétrica de los pacientes, sino que están fuertemente influenciados por factores geográficos y de transporte. Este fenómeno es particularmente marcado en el caso del Hospital Ludovica, cuya afluencia de pacientes desde el oeste y noroeste está claramente relacionada con la conectividad proporcionada por el Tren Roca.

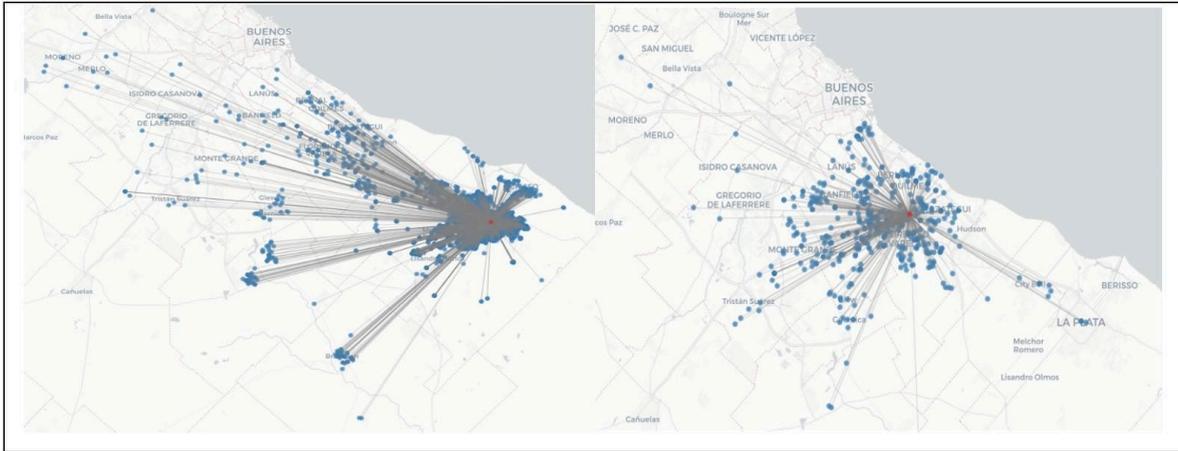
El uso de estimaciones de intensidad espacial mediante el método de Kernel (Figura 6) permite identificar “focos” de casos graves en distintas áreas del AMBA, destacando un incremento notable en la intensidad de casos severos en la región sur del AMBA para el año 2022 (en la Figura 6, notar que las escalas de intensidad son diferentes, siendo la de 2022 de aproximadamente el doble de la de 2019). La concentración de casos graves en áreas como Banfield y Florencio Varela, donde se ubica el Hospital El Cruce, refuerza la hipótesis de que la severidad de la enfermedad está estrechamente relacionada con la capacidad hospitalaria y la vulnerabilidad socioeconómica de las regiones circundantes. Sin embargo, también es necesario considerar el papel de la derivación de pacientes y cómo esto podría sesgar las estimaciones de intensidad espacial, ya que los casos más graves tienden a ser canalizados hacia hospitales de mayor complejidad.

El análisis del riesgo relativo, presentado en la Figura 7, revela diferencias espaciales en la severidad del VSR postpandemia en comparación con las infecciones leves o moderadas. Los resultados sugieren que en 2022 se extienden las regiones de mayor riesgo de casos graves de VSR en relación con las IRAB moderadas en relación con 2019, lo que subraya el posible efecto que podría haber tenido la pandemia de COVID-19 y las restricciones de circulación enfrentadas en 2020 y en parte del

2021.

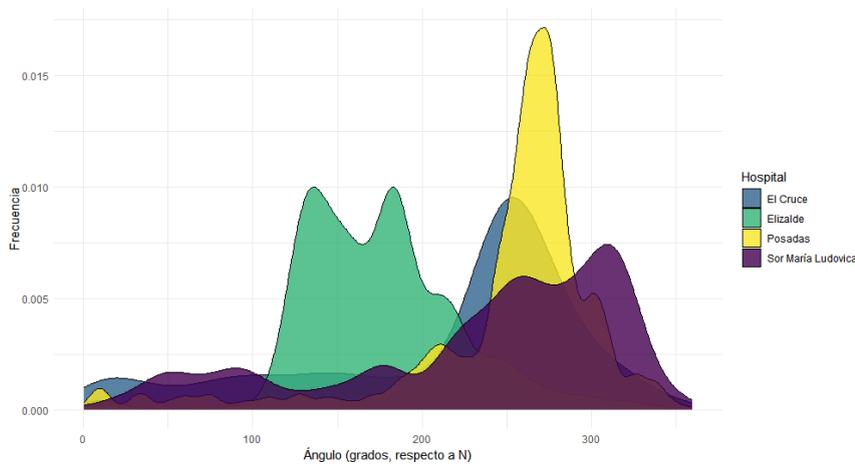
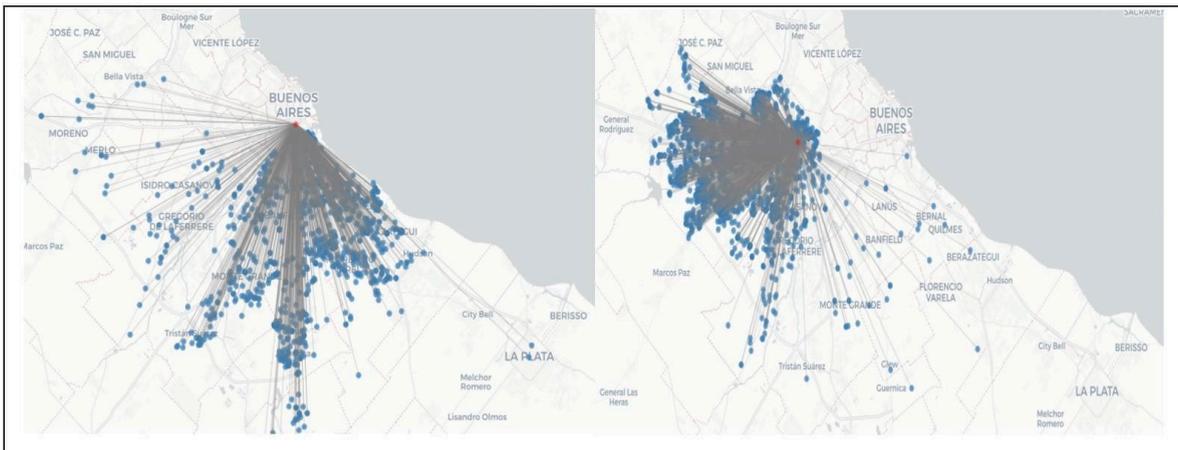
Ludovica (n = 4431)

El Cruce (n = 607)



Elizalde (n = 1737)

Posadas (n= 2515)



Ciudad de Buenos Aires, octubre 2024

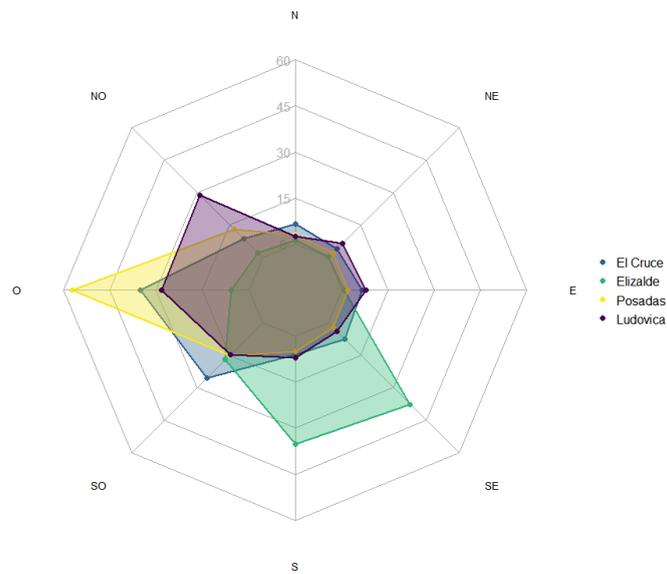


Figura 5. Gráfico superior. Flujos de visitas a los hospitales según los domicilios de los pacientes. Le sigue el histograma de los ángulos (respecto del norte) de los vectores de flujo, que unen cada hospital desde los domicilios de los pacientes que los visitan; en el gráfico inferior, esa información se complementa con un *radarplot*. En efecto, se observa que ningún hospital presenta una distribución razonablemente radial, sino que sus flujos tienen direcciones bien marcadas. Por ejemplo, el hospital Elizalde, en el sur de la CABA, presenta flujos desde el sur y el sureste del AMBA; mientras que el hospital Posadas recibe pacientes desde las direcciones oeste y suroeste. El hospital Ludovica, por su cercanía a la costa (este) y por la gran relación con el sistema de transporte ferroviario del Tren Roca, presenta una marcada afluencia desde la dirección oeste y noroeste. En todos los casos se observa que los hospitales tienen regiones de cobertura claramente definidas.

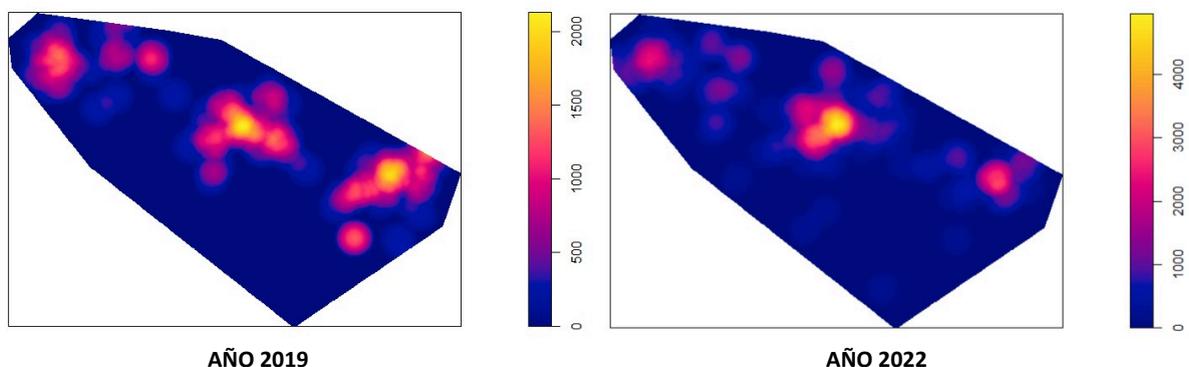


Figura 6. Estimación de la intensidad espacial de casos de VSR graves para los años 2019 y 2022, usando el mismo ancho de banda para ambas estimaciones; sin embargo, notar que las escalas de intensidad son diferentes, siendo la de 2022 de aproximadamente el doble de la de 2019. La intensidad estimada para los casos de enfermedad grave por sincicial en 2022 parece concentrarse en la región de cobertura del hospital El Cruce, en las cercanías de los partidos de Temperley, Banfield y Florencio Varela. Al respecto, cabe observar dos aspectos: por un lado, que el hospital El Cruce es un hospital de alta complejidad y derivación que, naturalmente, concentra la mayor cantidad de casos graves, razón por la cual es un foco de intensidad en ambas estimaciones (pre y post pandemia). Por otro, que la intensidad estimada en la región sur del AMBA (La Plata, casos del Ludovica) y en la región oeste del AMBA (Morón, casos del Posadas) es relativamente similar en ambos escenarios (pre y postpandemia), difiriendo más notablemente en la intensidad estimada para la región de cobertura del hospital El Cruce, lo que apoya la hipótesis del incremento en la severidad de la enfermedad postpandemia. Aunque la intensidad estimada para el año 2022 parece más concentrada en la región sur y oeste del AMBA respecto del 2019, cabe recordar que las escalas de intensidad de ambos gráficos son distintas, siendo la de 2022 aproximadamente del doble de la de 2019, lo que distorsiona la visualización inicial que se hace de estos gráficos.



Figura 7. Estimación del riesgo relativo de la enfermedad por VSR severo, en relación con la IRAB leve o moderada, para los años 2019 y 2022. Postpandemia, en el año 2022, puede observarse un mayor riesgo relativo de VSR severo respecto de las IRAB moderada en la región que comprende los municipios de Banfield, Temperley y Florencio Varela; todos casos que se concentran en el hospital de alta complejidad de El Cruce, sito en esta última localidad.

Por otro lado, observamos que el índice de vulnerabilidad sanitaria (IVS) parece capturar parcialmente las desigualdades en el acceso y calidad de los servicios de salud, lo que se observa en el impacto que tiene en la estimación de severidad controlando por edad (Figura 8). Sin embargo y como es sabido, el desafío en este tipo de análisis radica en desentrañar las correlaciones espurias entre las características espaciales y la gravedad de la enfermedad, algo que escapa al alcance del estudio dada la forma de recolección de los datos.

RESULTADOS Y DISCUSIÓN

Este estudio ofrece una mirada exploratoria sobre la dinámica espacial de las infecciones respiratorias en el AMBA, y en particular, para el caso del virus sincicial respiratorio. Si bien la exploración es clara en cuanto a exhibir posibles cambios en la dinámica espacial de la infección por VSR entre niños menores de cinco años con IRAB en el AMBA a partir de la pandemia de COVID-19, así como cambios en la gravedad de la enfermedad desde un enfoque espacial, y posibles correlaciones con variables socioeconómicas, no obstante, quedan varios desafíos metodológicos y conceptuales por explorar. La incorporación de modelos espaciotemporales que tengan en cuenta las características dinámicas de la movilidad poblacional y las capacidades hospitalarias podría mejorar la comprensión de los patrones observados. Además, sería valioso integrar modelos de simulación que exploren escenarios futuros bajo diferentes políticas de salud y transporte, lo que permitiría a los formuladores de políticas diseñar estrategias más efectivas para mitigar las infecciones respiratorias y reducir la carga sobre los servicios hospitalarios.

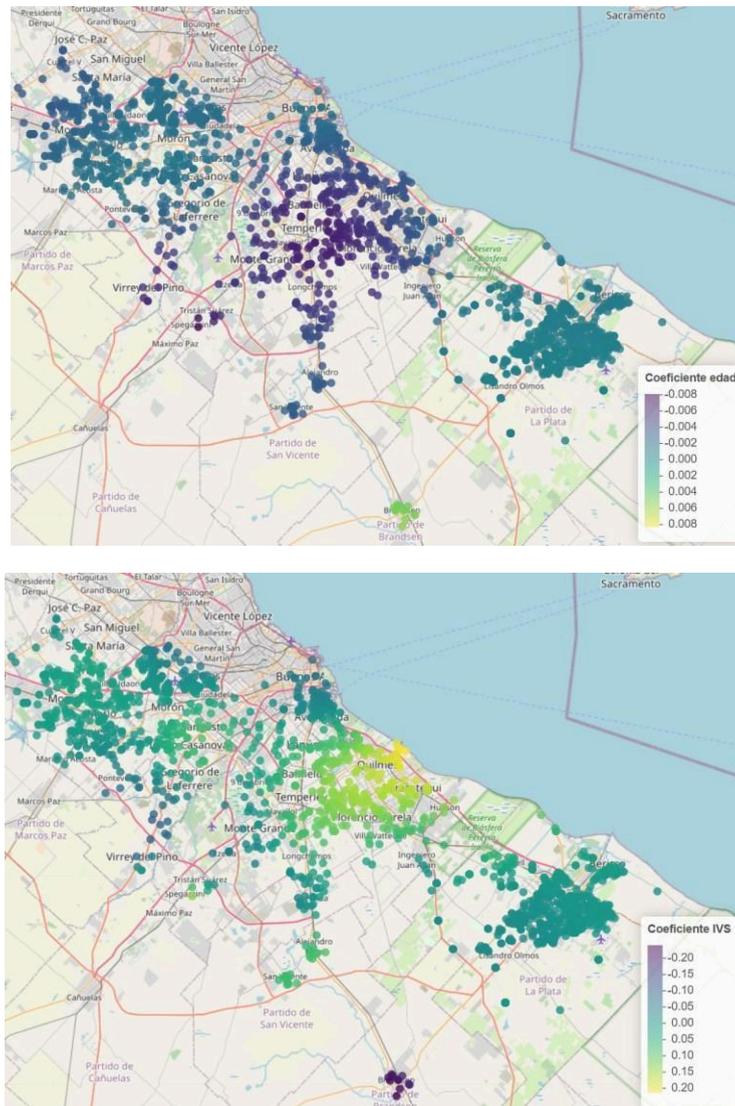


Figura 8. Coeficientes estimados en la regresión espacial ponderada para la predicción de la gravedad de la enfermedad por VSR en 2022, para las variables: edad en meses (arriba) e IVS (abajo). El coeficiente edad captura una relación esperable con la severidad: pacientes de menor edad son más vulnerables de desarrollar enfermedad grave. Los coeficientes estimados resultan negativos para casi todos los pacientes del estudio, con mayor impacto en aquellos con domicilios en la región suroeste del AMBA (partidos de Banfield, Temperley, Florencio Varela, Longchamps, Monte Grande, Virrey del Pino, entre otros; puntos con mayor intensidad de violeta en el gráfico superior). El IVS también tiene un impacto en la predicción de enfermedad severa: a mayor IVS, mayor es la gravedad de la enfermedad, y esto ocurre especialmente con los pacientes de la región sur, próxima a los partidos de Quilmes, Berazategui y cercanías en dirección oeste (puntos con mayor intensidad de violeta en el gráfico inferior). En ambos casos resalta lo que ocurre con los casos de Brandsen, que presentan un comportamiento llamativo

REFERENCIAS

- ACTUALIZACIÓN REGIONAL, INFLUENZA. Semana Epidemiológica 27 - Julio 21, 2021 - PS/OMS
Organización Panamericana de la Salud.
<https://www.paho.org/es/documentos/actualizacion-regional-influenza-semana-epidemiologica-27-julio-21-2021>.
- BIVAND, R.S., PEBESMA, E., GÓMEZ-RUBIO, V. (2013). Applied Spatial Data Analysis with R. Second Edition. New York: Springer.
- CABALLERO, M. T. ET AL. (2019). Mortality Associated with Acute Respiratory Infections Among Children at Home. *The Journal of Infectious Diseases* **219**, 358–364.
- CASALEGNO, J.-S. ET AL. (2021). Characteristics of the delayed respiratory syncytial virus epidemic, 2020/2021, Rhône Loire, France. *Eurosurveillance* **26**, 2100630.
- DIGGLE, P. J. (2014). Statistical Analysis of Spatial and Spatio-Temporal Point Patterns. Third Edition. UK: Taylor & Francis.
- GEOGHEGAN, S. ET AL. (2017). Mortality due to Respiratory Syncytial Virus. Burden and Risk Factors. *American Journal of Respiratory and Critical Care Medicine* **195**, 96–103.
- GOMEZ, G. B., MAHÉ, C., CHAVEZ, S. S. (2021). Uncertain effects of the pandemic on respiratory viruses. *Science*, Vol. 372, Issue 6546, pp. 1043-1044.
- HAAPANEN, M., RENKO, M., ARTAMA, M. & KUITUNEN, I. (2021). The impact of the lockdown and the re-opening of schools and day cares on the epidemiology of SARS-CoV-2 and other respiratory infections in children – A nationwide register study in Finland. *EClinicalMedicine* **34**.
- HODGSON, D., PEBODY, R., PANOVSKA-GRIFFITHS, J., BAGUELIN, M. & ATKINS, K. E. (2020). Evaluating the next generation of RSV intervention strategies: a mathematical modelling study and cost-effectiveness analysis. *BMC Medicine* **18**, 348.
- NAIR, H. ET AL. (2013). Global and regional burden of hospital admissions for severe acute lower respiratory infections in young children in 2010: a systematic analysis. *The Lancet* **381**, 1380–1390.
- POLACK, F. P. (2015). The changing landscape of respiratory syncytial virus. *Vaccine* **33**, 6473– 6478.
- SHI, T. ET AL. (2017). Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. *The Lancet* **390**, 946–958.

Índice por autor

- ABBA, 35
ABBIATI, 41, 46, 52, 55, 101
ABELLI BONARDI, 49, 196
ACETTA, 81, 105, 164
ACOSTA ALBARRACÍN, 33
AGUIRRE, 95
ALARCÓN, 129
ALEMÁN, 105
ALEN, 33, 159
ALESSO, 81, 105, 164
ALFARO, 62
AMARO, 143
ARCE, 44
ARES, 57
ARGÜELLES, 62
ARREDONDO, 32
ATTONATY, 142
BABINO, 23
BAHL, 61
BAILLIET, 62
BAJDA, 143
BALDO, 124
BALLEJO, 95
BALLIVAN, 114
BALZARINI, 42, 72, 74, 87, 186
BARBERÁ, 23
BEKINSCHTEIN, 98
BELDOMENICO, 100, 137
BENITEZ-VIEYRA, 57
BERNAUS, 78
BERNSTEIN, 145
BERRUETA, 114
BILENCA, 113
BÓBEDA, 157
BOCA, 34
BONADIE, 106
BORTOLOTTI, 39
BOSCOLO, 40, 98
BOVINA, 151
BRAMARDI, 48
BRASESCO, 69
BRAVI, 62
BRUNO, 39, 63, 65, 78, 108, 111, 116
BRUVERIS, 90
BRUZZONE, 86
BUSTOS, 104
CAGNONE F, 87
CAGNONE MF, 45, 150
CAMALET-LE NOBLE, 106, 107
CAMPANELLO, 102
CAMPERO, 156
CANESINI, 138
CANOVA, 35
CANTET, 17
CAO, 82, 169
CARBONELL-ASINS, 24
CARDELLI, 108
CARDINAL, 133
CARDOZO, 157
CAREAGA, 136
CARRIÓ, 75, 138
CASANOVES, 144
CASTELLANA, 114
CASTIGLIONE, 124
CASTRO, 99
CAVIA, 67
CENTURIÓN, 41
CERIANI, 142
CERVIGNI, 39
CHABBAL, 152
CHAPARRO, 69, 92
CHAUFAN, 123
CHIALINA, 117
CIAPPONI, 114
CIPRIOTTI, 81, 129, 164
COALOVA, 123
CODESIDO, 113
CORRAL, 11, 43, 58
CORVALÁN, 89
COSTA TÁRTARA, 44, 48
COY, 83
CRISTIANO, 45, 86, 102, 126, 150
CUERVO, 137
CUESTA, 47, 191
CUESTA NÚÑEZ, 85
CUETO, 85, 127
CURIA, 50
CURTI, 48
CÓRDOBA, 42, 74, 87, 111, 186
DAMIANI QUIROZ, 86
DAMICO, 55
DE LELLIS, 45, 102
DE LOYOLA, 146
DE MARTINO, 32
DE ROIA, 133
DEL DUCA, 94, 200
DEL MÉDICO, 51
DEMETRIO, 61
DI PASQUALE, 55
DI RIENZO, 17, 144
DÍAZ VILLA, 45, 86, 126, 150,
DIPIERRI, 62
DOPAZO, 62
DRAMIS, 28
DRAQUE, 89
DURE RUIZ, 131
DUS SANTOS, 36
DVORKIN, 33, 159
ELIAS-COSTA, 124
ENRIQUEZ, 133
ERNETA, 46
FAZIO, 136
FEIJÓO, 54
FELLET, 99
FERNÁNDEZ E, 62
FERNÁNDEZ GJ, 43, 131
FERNÁNDEZ JC, 49, 196
FERNÁNDEZ MS, 85, 145, 153
FERNÁNDEZ P, 138
FERREYRA A, 66
FERREYRA S, 41, 58
FERRUFINO, 36



FIGUEROA, 62, 110	IGLESIAS J, 63, 116	MENSCH, 155
FILIPIGH, 63, 78	IGLESIAS PP, 124	MERGEL, 52
FILLOY, 59, 96, 148, 151	INCHAUSTI, 24	MICCOLI, 41, 55, 58
FIORE, 65	IORIO, 41, 58	MIGLIETTA, 95
FLAIBANI, 145	IZZO, 47, 191	MILESI, 82, 169
FLESIA, 64	JUANEDA ALLENDE, 66	MILICIA, 52
FLORES YAPU, 64	JULIANO, 41, 58	MINARDI, 35
FLOREZ, 84	JUNGES, 100	MINOLI, 117
FORTE, 18	KNACK DE ALMEIDA, 83	MIÑO, 28
FRANK, 77, 79, 174, 180	KREIMAN, 155	MIQUEL, 56
FRASCA, 117	KUPERMAN, 37	MIRAGALLA, 95
FURLAN, 100	LA SALA, 129	MIRETTI, 62, 82, 169
FURMAN, 98	LAC PRUGENT, 48, 143	MODINI, 75
GALARZA, 41, 58	LAFFITTE, 49, 140, 196	MOLINARI, 92
GALIOTTI, 89	LAJE, 98	MONTENEGRO BRUSOTTI, 106, 107
GAMARNIK, 95	LANDAJO, 77, 174	MONTES, 32
GANDINI, 118	LAPRIDA, 91	MORALES, 70, 111, 121, 135
GARCÍA CHAPORI, 91	LAURENZANO, 49, 196	MORELLI, 89
GARCIA SELEME, 87, 111, 186	LAVALLE, 49, 50, 51, 140, 196	MUCHIUT, 66
GARCÍA TOURN, 49, 196	LIBRYK, 149	MUNILLA, 146, 147
GARIBALDI, 37	LILLO, 49, 196	MURILLO, 75
GARZÓN, 67	LLERA, 14, 89, 140	MUÑIZ SAAVEDRA, 196
GASPE, 133	LO PRESTI, 78	MUÑOZ MORALES, 70
GÉNERO, 46	LÓPEZ-MIRANDA, 95	NATAL, 92
GERMAN, 66	LÓPEZ DV, 78, 95	NATALE, 61
GIMÉNEZ G, 88	LÓPEZ G, 101	NAZRALA, 89
GIMÉNEZ L, 71, 108, 109, 152	LÓPEZ MJ, 90	NIKLISON, 113
GOLATO, 54	LÓPEZ S, 28	NUÑEZ, 99, 106, 107
GOLDIN, 40, 98	LÓPEZ DE CASENAVE, 82, 169	OCAMPO, 84
GOLDSTEIN, 126	LUISI, 62	OJEDA, 133
GONZÁLES, 62	MACHIAVELLI, 19	OTERO, 53, 73
GONZÁLEZ AUDINO, 142	MAGGI, 100	PACCIORETTI, 42, 87, 111, 186, 190
GONZÁLEZ F, 36	MALASPINA, 88	PALAVECINO, 137
GONZÁLEZ R, 84	MANSUR, 66	PALMERIO, 132
GONZÁLEZ GARELLO, 136, 145	MARFETAN, 114	PARADA, 28, 33, 94, 159, 200
GOROSITO, 119	MARRO, 95	PARDINI, 145
GUERRERO, 58, 69	MARTINENGO, 89	PARONETTO, 45, 86, 126
GUTIERREZ, 54, 156	MARTINEZ, 64	PARRA, 81, 164
HAIQUE, 88	MARTÍN, 91	PASQUALE, 95
HARTIG, 18	MARTÍNEZ, 89, 147	PAZ, 40, 62
HELMAN, 145, 153	MARZIALI BERMÚDEZ, 119	PELETTIERI, 119
HERMIDA, 40	MASSONI, 132	PELOSO, 99
HIDALGO, 152	MAURE, 89	PEREYRA, 124
HUAYLLA, 37	MAYR, 20	PÉREZ ADRIANA, 90, 127, 153
IENO, 19	MEJÍA, 85	
	MENDOZA, 68, 157	



PÉREZ ALEJANDRO, 129	RUZ, 28	VULLO, 82, 169
PEROTTO, 65	SABINO, 81, 108, 110, 164	VÁZQUEZ CAÑAS, 133
PETRONAVE, 99	SALGADO COSTA, 61	WEHRENDT, 58
PEÑA, 54, 70, 101, 116	SÁNCHEZ-LORIA, 62	ZAMBIANCHI, 106, 107
PEÑA MALAVERA, 54, 70	SÁNCHEZ FIGUEROA, 110	ZAMORA RUEDA, 54
PEÑAS BALLESTEROS, 116	SANTADINO, 118	ZAMUDIO, 71
PICHLER, 24	SANTOS, 36, 61	ZERBATO, 75
PISTONESI, 64	SCHIAFFINO, 59	ZULIANI, 140
PITA, 123	SCHILMAN, 155	
PIZARRO, 75	SCREPNIK, 71	
POKLEPOVICH, 62	SHALOM, 40	
POL, 82, 169	SILVA, 95	
PONCE DE LEÓN, 152	SMREKAR, 66	
PONCET, 95	SORIA, 62	
PORRINI, 69	SOSA, 89	
PORTA, 56	SOSA ESCUDERO, 21	
PRATTA, 44, 51	SOTO E, 56	
PUHL, 139	SOTO I, 56	
PUPPO, 34	STOKER, 138	
RACCA, 137	STRATICO, 53, 73	
RAMALLO, 62	SUÁREZ F, 63, 74, 111	
RAMOS, 138, 151	SUÁREZ JC, 144	
RAS, 77, 174	SVENDSEN, 84	
RELA, 98	SZRETTER NOSTE, 26	
RIBERO, 59	TABLADA, 111	
RICARDO, 95	TAUBER, 28, 110	
RICHMOND, 53, 73	TEDESCO, 149	
RISSO, 25	TKACH, 96	
RIVERA, 49, 196	TOGNETTI, 150	
ROBALINO, 137	TOLOSA, 44	
ROBLES, 35	TOPAYAN, 46, 147	
RODRIGUEZ-MOZAZ, 61	TORRES, 49, 196	
RODRÍGUEZ A, 66	TREZZA-NEUMAYER, 133	
RODRÍGUEZ NL, 104	TRONCOSO, 49, 196	
RODRÍGUEZ PE, 95	TURJANSKI, 28, 83, 153	
RODRÍGUEZ AZCONA, 157	UGOLINI, 49, 196	
RODRÍGUEZ NUÑEZ, 72	UNZAGA, 153	
ROJAS, 144	VACCARO, 148	
ROMERO, 84, 104	VAIRA, 75, 92, 138	
RONDANINI, 139	VALDORA, 123	
ROSSI, 95	VELEZ RUBIO, 64	
ROVEGNO, 41, 55, 101	VELÁZQUEZ LÓPEZ, 78	
ROZENBLIT, 127, 145	VERCELLINI, 32	
RUBIO, 88, 108, 109	VIGNEAU, 139	
RUGGIERO, 149	VILLARREAL, 56	
RUIZ DÍAZ, 66	VITELLESCHI, 51	
RUMINOT, 108, 109	VOVCHUK, 136	
RUÍZ, 54	VRDOLJAK, 57	



XXVIII Reunión Científica del Grupo Argentino de Bioestadística



La educación y la ciencia fortalecen sociedades más equitativas, soberanas y democráticas

Ciudad de Buenos Aires, octubre 2024