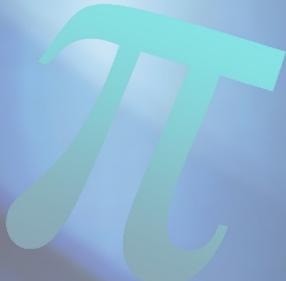


LIBRO DE RESÚMENES



Congreso Interamericano de Estadística

17 al 20 de octubre 2017
Rosario, Argentina





Congreso Interamericano de Estadística

PROGRAMA

XLV Coloquio Argentino de Estadística de la
Sociedad Argentina de Estadística
XXII Reunión Científica del Grupo Argentino de Biometría
Jornadas del Instituto Interamericano de Estadística

17 al 20 de octubre de 2017

Rosario, Argentina

www.cie2017.s-a-e.org.ar/congreso.php

Organizado por:



Apellido, Nombre

Libro de Resúmenes Congreso Interamericano de Estadística. - 1a ed.

Rosario: Editora, 2017. X p. ; 21x16 cm.

ISBN 000-000-0000-00-0

1. Estadísticas. I. Título.

CDD 310.4

Fecha de catalogación: 01/01/2017

© 2017 Congreso Interamericano de Estadística
ISBN 000-000-0000-00-0

Publicado por:

Sociedad Argentina de Estadística

Av. Corrientes 4264, 9º piso A

C1195AAO Ciudad Autónoma de Buenos Aires

Compilado por:

Nombre Apellido

email@gmail.com

Rosario - Argentina

Organizadores del Congreso:

Sociedad Argentina de Estadística (SAE)

Grupo Argentino de Biometría (GAB)

Instituto Interamericano de Estadística (IASI)

Impresión:

Poner el nombre de la editora si se imprime

Fecha de impresión: Octubre de 2017

El Congreso Interamericano de Estadística
cuenta con las siguientes declaraciones de interés:

- Declaración de Interés del Instituto Nacional de Estadística y Censos
- Declaración de Interés Educativo del Ministerio de Educación de la Provincia de Santa Fe
- Declaración de Interés del Instituto Provincial de Estadística y Censos
- Declaración de Interés de la Cámara de Diputados de la Provincia de Santa Fe
- Declaración de Interés del Concejo Municipal de la Municipalidad de Rosario
- Declaración de Interés Académico del Consejo Superior de la Universidad Nacional de Rosario
- Declaración de Interés Académico del Sr. Rector de la Universidad Nacional de Rosario
- Declaración de Interés Académico del Consejo Directivo de la Facultad de Ciencias Económicas y Estadística
- Declaración de Interés Académico de la Sra. Decana de la Facultad de Ciencias Económicas y Estadística
- Declaración de Interés Institucional del Sr. Decano de la Facultad de Ciencias Agrarias
- Declaración de Interés de la Dirección General de Estadística de la Municipalidad de Rosario



Acompañamiento económico de:

- Presidente de la Cámara de Diputados de la Provincia de Santa Fe, Diputado Provincial Antonio Bonfatti.
- Diputado Provincial Joaquín Blanco.
- Concejal de la ciudad de Rosario Verónica Irizar

Contenido

| | |
|---|------------|
| Preliminares | 9 |
| Autoridades | 11 |
| Antecedentes | 13 |
| Comunicaciones orales | 15 |
| Control de procesos | 17 |
| Datos categóricos | 18 |
| Datos de duración | 19 |
| Datos faltantes | 21 |
| Diseño de experimentos | 22 |
| Estadística espacial | 24 |
| Inferencia estadística | 27 |
| Métodos bayesianos | 29 |
| Métodos multivariados | 33 |
| Métodos no paramétricos | 36 |
| Minería de datos | 37 |
| Modelos de regresión | 39 |
| Muestreo | 43 |
| Probabilidad y procesos estocásticos | 48 |
| Series de tiempo | 49 |
| Teoría sobre distribuciones de probabilidad | 51 |
| Otras categorías metodológicas | 52 |
| Pósters | 63 |
| Datos categóricos | 65 |
| Datos de duración | 67 |
| Datos faltantes | 69 |
| Diseño de experimentos | 71 |
| Estadística espacial | 76 |
| Inferencia estadística | 81 |
| Métodos bayesianos | 87 |
| Métodos multivariados | 91 |
| Métodos no paramétricos | 105 |
| Métodos robustos | 106 |
| Minería de datos | 107 |
| Modelos de regresión | 109 |
| Muestreo | 130 |
| Probabilidad y procesos estocásticos | 133 |
| Probabilidades | 134 |
| Series de tiempo | 135 |
| Otras categorías metodológicas | 137 |
| Índice de autores | 160 |

PRELIMINARES

Autoridades

Presidente Cristina Cuesta. Facultad de Ciencias Económicas y Estadística. UNR

Vicepresidente Alberto Trevizán. Facultad de Ciencias Agrarias. UNR

Miembros Docentes de la Facultad de Ciencias Económicas y Estadística. UNR

Docentes de la Facultad de Ciencias Agrarias. UNR

Docentes de la Facultad de Ciencias Exactas, Ingeniería y Agrimensura.
UNR

Comité Científico

Arce, Osvaldo (UNT)

Beltran, Celina (UNR)

Barbona, Ivana (UNR)

Blaconá, María Teresa (UNR)

Boggio, Gabriela (UNR)

Bussi, Javier (UNR)

Casperri, María Teresa (UBA)

Castro Kuriss, Claudia (ITBA)

Cendoya, María Gabriela (UNMdP)

Charre de Trabuchi, Clyde (IASI)

Cosolito, Patricia (UNR)

Cuesta, Cristina Beatriz (UNR)

De Alba, Enrique (IASI)

Diaz, María del Pilar (UNC)

Di Blasi, Angela (UNCu)

Di Rienzo, Julio (UNC)

García, María del Carmen (UNR)

Giménez, Laura (UNNE)

Hachuel, Leticia (UNR)

Kelmansky, Diana (UBA)

Llera, Joaquín (UNC)

Marí, Gonzalo (UNR)

MacDonald, Alphonse (IASI)

Mendez, Fernanda (UNR)

Pagura, José A. (UNR)

Quaglino, Marta (UNR)

Quintslr, Marcia (IASI)

Ricci, Lila (UNMdP)

Rosa, Ernesto (UNTREF)

Trevizán, Alberto (UNR)

Urrutia, María Inés (UNLP)

Antecedentes

De la Escuela de Estadística de la UNR

La carrera de Estadística de la Facultad de Ciencias Económicas y Estadística de la UNR fue creada en el año 1948 por iniciativa del Profesor Carlos Eugenio Dieulefait en la entonces Facultad de Ciencias Económicas de la Universidad Nacional del Litoral. Este evento significó un paso trascendental para la estadística de nuestro país ya que fue la primera carrera en el área de Latinoamérica.

A lo largo de su historia egresaron numerosos profesionales, de los cuales muchos se distinguieron en el país y en el extranjero desempeñándose en el ámbito académico, reparticiones públicas y privadas, en distintas áreas tales como estadísticas oficiales, ciencias económicas, biología y salud, agronomía, industria, demografía, entre otras.

Actualmente la carrera de grado, junto a la Maestría en Estadística Aplicada, creada en el año 2002, y el Doctorado en Estadística, creado en el año 2014, ambos de la UNR, forman parte de uno de los principales centros del país de reconocido prestigio, en formación de recursos humanos en el área. Cabe señalar que la Escuela de Estadística posee una extensa trayectoria en investigación desarrollada en los institutos de investigaciones. Las tareas en investigación no sólo han enriquecido la formación docente, sino que también motivan la participación de los alumnos en las mismas.

Las actividades de transferencia, que se desarrollan a través de convenios firmados entre la Escuela de Estadística y organismos públicos y privados, permiten a docentes y estudiantes la vinculación con el medio y la resolución de problemas reales. Estas características y la evolución de los planes de estudio han permitido lograr un perfil de graduado de fácil inserción laboral en distintos ámbitos del país y del exterior, con posibilidades, además, de realizar con éxito estudios de posgrado en universidades nacionales y extranjeras.

Del Instituto Interamericano de Estadística

El Instituto Interamericano de Estadística (IASI) fue fundado en el año 1940, a iniciativa de un grupo de miembros del Instituto Internacional de Estadística (ISI), interesados en mantener y desarrollar en la región americana actividades y programas que a nivel mundial venía llevando a cabo el ISI, y que se habían interrumpido a consecuencia de la II Guerra Mundial.

El Instituto Interamericano de Estadística (IASI) es una organización profesional cuyo propósito es promover el desarrollo de la estadística en la región americana. Persigue los siguientes objetivos:

- (a) desarrollo y fortalecimiento de la profesión estadística;
- (b) promoción y divulgación de avances en teoría y métodos estadísticos;
- (c) perfeccionamiento de la metodología en la producción de estadísticas, tanto gubernamentales como no gubernamentales;
- (d) promoción de medidas que tiendan a mejorar la comparabilidad y el aprovechamiento de las estadísticas económicas y sociales entre las naciones de la región; y
- (e) colaboración con las organizaciones nacionales e internacionales en actividades orientadas al mejoramiento de la estadística en la región.

De la Sociedad Argentina de Estadística

La Sociedad Argentina de Estadística (SAE) es un organismo técnico - científico que promueve el desarrollo de la Estadística en el país. Fue creado en junio de 1952 y entre sus objetivos se encuentran: agrupar y vincular entre sí a todas las personas relacionadas con la Ciencia Estadística y con sus aplicaciones; fomentar y promover el estudio, investigación, desarrollo y perfeccionamiento de la Estadística y disciplinas conexas, mediante reuniones científicas y técnicas, cursos,

becas, premios, concursos, publicaciones sin fines de lucro y otros medios adecuados; contribuir al mejoramiento de la enseñanza de la Estadística y mantener relaciones con instituciones afines, nacionales y extranjeras.

Cada año la SAE organiza Coloquios que intercaladamente comparte con otros países de Latinoamérica.

Del Grupo Argentino de Biometría

El Grupo Argentino de Biometría (GAB) nace en 1995, con el objeto de relacionar personas que investigan y/o enseñan aspectos de la estadística aplicada a la biología, promoviendo la investigación, la docencia y la difusión de estos conocimientos a través de diversas actividades científicas.

Sus objetivos principales son: fortalecer y enriquecer los vínculos y propósitos que dieron origen al Grupo Argentino de Biometría (GAB); proveer un ámbito para el intercambio y discusión de ideas motivadoras e inspiradoras para la comunidad de biometristas; hacer público y difundir los resultados del trabajo de los biometrías argentinos; promover y/o fortalecer la generación de vínculos inter instituciones; fortalecer la formación de recursos humanos y la actividad de los biometristas en la Argentina.

COMUNICACIONES ORALES

Control de procesos

GRÁFICOS DE CONTROL DE PROCESOS DE ALTA CALIDAD MULTI-ATRIBUTO MEDIANTE LA DISTRIBUCION MULTINOMIAL

ANDREA RIGHETTI¹, SILVIA JOEKES¹ y MARCELO SMREKAR²

¹ INSTITUTO DE ESTADÍSTICA Y DEMOGRAFÍA, U.N.C. ²LAB. DE INGENIERÍA Y MANTENIMIENTO INDUSTRIAL, U.N.C.

analizamos@yahoo.com.ar

El mejoramiento continuo es esencial para la supervivencia y el crecimiento de las empresas que requieren de una permanente reducción en el nivel de no conformidades de sus procesos. Para lograr altos estándares de calidad, las empresas necesitan de la aplicación de programas de gestión y de procedimientos estadísticos adecuados. En este aspecto, el Control Estadístico de Procesos (CEP) sigue desempeñando un rol fundamental en el esfuerzo hacia el mejoramiento continuo. Controlar y asegurar la conformidad del producto con las especificaciones en niveles cercanos a cero defectos, requiere de procedimientos estadísticos, tales como gráficos de control cada vez más complejos. A pesar de que en la última década ha habido un creciente interés por el desarrollo de gráficos de control de procesos de alta calidad, existen a la fecha pocos procedimientos para controlar procesos cuya calidad está determinada por la presencia o ausencia de varias características no medibles, es decir, procesos en los que cada ítem es susceptible de ser clasificado en dos o más categorías simultáneamente. Un método de generalización de la distribución binomial es la distribución multinomial que puede ser empleada en la construcción de un gráfico de atributo multivariado. Este gráfico llamado gráfico de atributo multinomial (MAC) tiene en cuenta la correlación definiendo adecuadamente las clases de calidad de un producto. En este trabajo se presenta una aplicación considerando dos características de calidad binomiales. El procedimiento consiste en adicionar una nueva variable que surge del conteo de no conformidades para ambas variables simultáneamente más una categoría adicional que considera los ítems conformes. Para ello se aplica la estadística, se determina el atributo responsable de la salida de control del proceso mediante una estadística Z y se evalúa su rendimiento mediante la longitud media de corrida (ARL).

Palabras claves: procesos de alta cal; gráficos multi-atrib; gráfico de control

GRÁFICOS DE CONTROL MULTIVARIADOS PARA PROCESOS DE ATRIBUTOS DE ALTA CALIDAD

SILVIA JOEKES¹, ANDREA RIGHETTI¹ y MARCELO SMREKAR²

¹ INSTITUTO DE ESTADÍSTICA Y DEMOGRAFÍA, U.N.C. ²LAB. DE INGENIERÍA Y MANTENIMIENTO INDUSTRIAL, U.N.C.

silviajks@gmail.com

El estudio de múltiples características es importante en los procesos de alta calidad. En las cercanías de cero no conformidades la mayoría de los ítems serán conformes, especialmente cuando se considera una sola característica crítica. En el caso de considerar varias características, los productos pueden ser mejor diferenciados y las oportunidades de mejoras suelen ser más fácilmente identificables, permitiendo obtener mayor información acerca de un producto o proceso. Los gráficos de control de atributos son utilizados para el monitoreo de procesos donde las características de calidad no pueden medirse en una escala continua. Esta situación se observa en procesos industriales, procesos relacionados con la salud, y en procesos de servicios, entre otros. Muchos de estos ejemplos

implican el monitoreo simultáneo de múltiples atributos lo que conduce a la aplicación de métodos de control de calidad multivariados que han mostrado ser más adecuados que el uso simultáneo de múltiples procedimientos univariados. En este estudio se presenta en primer lugar una breve revisión de la metodología existente para procesos multivariados de alta calidad. A continuación se muestra la metodología y aplicación de un gráfico de control multivariado, el gráfico Mnp que surgió como una extensión de los gráficos np univariados de Shewhart. Este gráfico emplea una estadística que se obtiene como la suma ponderada de los conteos de unidades no conformes para todas las características de calidad, considerando además la correlación entre atributos. Al mismo tiempo, se muestra un procedimiento sencillo para identificar la presencia de causas asignables ante una señal de fuera de control. Por último se realiza la comparación del gráfico Mnp con los gráficos np univariados, que son los más comúnmente empleados.

Palabras claves: procesos de alta cal; gráficos multivariad; gráfico de control m

Datos categóricos

DIAGNÓSTICO DE MODELOS DE MÚLTIPLES FACETAS DE RASCH EN ESTUDIOS OBSERVACIONALES

ROBERTO DANIEL CÁCERES BAUER

DEPARTAMENTO DE EDUCACIÓN MÉDICA, FACULTAD DE MEDICINA, UNIVERSIDAD DE LA REPÚBLICA, URUGUAY; UNIDAD DE ESTADÍSTICA, ESCUELA DE NUTRICIÓN, ÁREA INVESTIGACIÓN, UNIVERSIDAD DE LA REPÚBLICA, URUGUAY..

rcaceresb@gmail.com

El modelo de múltiples facetas de Rasch (MFRM) tiene aplicaciones en diversas áreas (Bond y Fox, 2015; Engelhard, 2013; Linacre, 2015). MFRM permite modelar la respuesta a ítems de una escala como un fenómeno probabilístico influenciado por elementos de varias facetas. Cada faceta corresponde a factores que afectan la probabilidad de la respuesta descripta mediante una variable ordinal. Un aspecto clave en la aplicación del MFRM es la conexión entre los elementos de las facetas. Este aspecto del diseño afecta la exactitud y precisión de las medidas y la aplicación de los procedimientos de diagnóstico del modelo. En estudios experimentales la conexión es definida a priori. No obstante, en estudios observacionales la conexión emergente puede ser compleja y deficitaria. Por tanto, se requieren métodos especializados de diagnóstico con los que actualmente no se cuenta. Los objetivos de este trabajo son: (a) caracterizar la conexión emergente en un estudio observacional; (b) comparar un método de diagnóstico corrientemente usado (Linacre, 2015), con un método alternativo. Los datos considerados provienen de un cuestionario de evaluación estudiantil de la calidad en la enseñanza empleada en un estudio observacional, con 3243 estudiantes, 98 docentes, 10 actividades de enseñanza, en una carrera de médico. El modelo MFRM incluyó tres facetas de medición (docentes, estudiantes e ítems). Para caracterizar la conexión emergente se aplicó un modelo de gráfica no dirigida (Trudeau, 2013) y se determinó el tipo de red de evaluación (Engelhard, 1996). En la comparación de los métodos de diagnóstico se consideró: la estabilización de las mediciones; el ajuste global; y las medidas de invarianza. La conexión en la faceta docentes resultó mayor que la observada en la faceta estudiantes. La red de evaluación encontrada resultó ser una red de evaluación conectada incompleta e incluyó conjuntos de elementos que son claves para la conexión de la red. Los criterios de comparación considerados sugieren que el método propuesto funciona mejor.

Palabras claves: modelos de múltiples; redes de evaluación; diagnóstico de model; estudios observacion; gráficas no dirigida; conexión

PROBABILIDAD DE MOROSIDAD EN LOS CRÉDITOS DE CONSUMO DE UNA COOPERATIVA DE AHORRO Y CRÉDITO DE PARAGUAY

MARIO DAMIÁN VÁZQUEZ¹ y MARÍA CRISTINA MARTÍN²

¹DIRECCIÓN GENERAL DE INVESTIGACIÓN Y ESTUDIOS ESPECIALES. UNIVERSIDAD NACIONAL DE VILLARRICA DEL ESPÍRITU SANTO. PARAGUAY. ²DEPARTAMENTO DE MATEMÁTICA. UNIVERSIDAD NACIONAL DEL SUR. ARGENTINA.

mscientiae@gmail.com

El objetivo del trabajo de investigación es estimar la morosidad de los socios de una cooperativa de ahorro y crédito de Paraguay al solicitar un crédito de consumo. Analizando la base de datos de la cooperativa y al aplicar la Regresión Logística para respuesta binaria con la técnica Stepwise se seleccionan las variables que componen el modelo final. Para la interpretación del modelo, lo que se analiza son los cocientes de ventaja (Odds Ratio), dada por la expresión $OR=e^{\beta_k}$, en términos probabilísticos, refleja el cambio marginal de pasar de la categoría de referencia a otra categoría de esta variable. Si la OR resulta inferior a la unidad, se calcula su inversa OR^{-1} para una mejor interpretación, siendo ahora la ventaja a favor de la categoría de referencia. Clasificando a los socios con probabilidad de ser morosos mayor a 0,16 como socios morosos, se obtienen los índices para medir la bondad de ajuste del modelo final estimado por Regresión Logística, siendo la Precisión igual a 70,55%, la Especificidad igual a 70,54% y la Sensibilidad igual a 70,59%, por lo que la capacidad predictiva del modelo matemático propuesto es aceptable. Similarmente, el valor 0,721 del área bajo la curva ROC, permite considerar que el modelo tiene capacidad de discriminación aceptable. Se concluye que hay evidencias estadísticas significativas para afirmar que el plazo solicitado en el crédito de consumo en categorías, la antigüedad en la cooperativa en categorías, el estado civil del socio y el monto solicitado en el crédito, estiman la probabilidad de morosidad de los socios. Los resultados son relevantes para las políticas de utilización de modelos matemáticos en la estimación de la morosidad en los créditos de consumo en la cooperativa.

Palabras claves: regresión logística; técnica stepwise; odds ratio; curva roc; morosidad; cooperativa

Datos de duración

BIVARIATE GENERALIZED MODIFIED WEIBULL DISTRIBUTION FOR LIFETIME MODELING WITH A CURE FRACTION

MARCOS VINICIUS DE OLIVEIRA PERES, JORGE ALBERTO ACHCAR y EDSON ZANGIACOMI MARTINEZ

DEPARTMENT OF SOCIAL MEDICINE, UNIVERSITY OF SÃO PAULO, RIBEIRÃO PRETO MEDICAL SCHOOL, BRAZIL.

mvperes1991@usp.br

In medical studies for lifetime modeling we can find situations where a fraction of individuals will not experience the event of interest. In this case the traditional approaches for survival analysis are not suitable, since it is assumed that all individuals under study are susceptible to the event of interest. In such cases, the survival regression models in the presence of a cure fraction can be more

appropriate to fit the data. Other aspect of survival analysis that should be given attention in the medical applications is the presence of multivariate lifetimes, in particular bivariate lifetimes. We have as examples the time until a treated eye fails or the time to recurrence of each breast after cancer treatment. Survival copula functions have become widely used for modelling multivariate survival data. Copulas are functions that connect univariate distribution functions to form a multivariate distribution. For more information, read Nelson (2006) or Balakrishnan and Lai (2009). In this work we consider a bivariate model for survival data based on the four-parameter generalized modifield Weibull distribution (Carrasco et al., 2008) in the presence of a cure fraction and right-censored data. A Bayesian approach was applied and the parameter estimation was based on Markov Chain Monte Carlo (MCMC) simulation methods. We consider two types of copula functions, the Farlie-Gumbel-Morgenstern (FGM) and Clayton copulas. We also consider independence between the lifetimes. In addition, it was compared the situations where the cure fractions for lifetimes are dependent or independent of one another. Comparison between models from different formulations was assessed by the deviance information criteria (DIC) and visual comparisons between the predicted marginal survival functions based on the proposed models and the respective Kaplan-Meier estimates. To illustrate the applicability of the model, we consider a real data set related to an invasive cervical cancer study. In this application, we conclude that the model based on the Clayton copula and which assumes dependence between the parameters related to the cure fractions is the most appropriate to the data.

Palabras claves: copula functions; survival analysis; bayesian analysis

ESTUDIO SOBRE LA DEMORA EN LA ELECCIÓN DE CARRERA EN UNA FACULTAD DE ROSARIO MEDIANTE ANALISIS DE SUPERVIVENCIA PARA TIEMPOS AGRUPADOS

JULIA ANGELINI y GABRIELA BOGGIO

INSTITUTO DE INVESTIGACIONES TEÓRICAS Y APLICADAS DE LA ESCUELA DE ESTADÍSTICA, FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA, UNIVERSIDAD NACIONAL DE ROSARIO..

jangelini.93@hotmail.com

Uno de los problemas que enfrentan las instituciones de Educación Superior relacionado con el rendimiento académico se refiere a la excesiva demora para alcanzar el título. En particular, se ha percibido en la Facultad de Ciencias Económicas y Estadística de la Universidad Nacional de Rosario un retraso en los tiempos previstos por los planes de estudio para cumplir una primera meta relacionada con la posibilidad de elegir carrera. El Plan de Estudios de las carreras de Cs. Económicas de esta Facultad comienza con un Ciclo Introductorio Común y la elección de la carrera es posible previa aprobación de cuatro de las seis asignaturas que conforman ese ciclo. Se ha notado una demora importante en cumplir este requisito, por lo que interesa investigar cuáles son los determinantes que tienen impacto directo sobre ella. Para dar respuesta a este objetivo, se recurre a técnicas específicas de análisis de datos de supervivencia para tiempos discretos o agrupados, debido a que, si bien el cumplimiento del requisito para poder elegir carrera se puede dar en diferentes momentos durante el año académico, en este trabajo se utiliza la información disponible al inicio de cada año académico, es decir, la demora se mide en cantidad de años. El enfoque elegido consiste en tratar cada tiempo de supervivencia individual como un conjunto de observaciones dicotómicas que consideran si un alumno presenta o no el evento en cada año académico hasta que experimente el evento o sea censurado. Esta reestructuración de los datos permite ajustar un modelo lineal generalizado para respuesta binaria eligiendo, en este caso, el enlace “log-log del complemento”, de manera de interpretar los coeficientes en términos de razones de hazards como en el clásico modelo de Cox. Este enfoque permite, además, la inclusión de covariables que dependan del tiempo sin

dificultades y la obtención de una medida del efecto de que el evento ocurra en un determinado tiempo en comparación con que ocurra en otro tiempo. Los resultados hallados muestran que los estudiantes que tienen mayor demora en el cumplimiento de los requisitos para elegir carrera son los varones, los egresados de una escuela secundaria pública, los estudiantes que no trabajan, los que tienen un promedio menor a seis y aquéllos cuyo padre tiene un nivel educativo bajo. Por último, presentan mayor demora los que han terminado la educación secundaria antes de 2007, lo que se ve potenciado si además las madres trabajan.

Palabras claves: censura por intervalo; modelos lineales gen; razón de hazards

INFERENCIA Y BONDAD DE AJUSTE PARA DATOS CON CENSURAS HÍBRIDA Y PROGRESIVA

CLAUDIA CASTRO KURISS¹ y VÍCTOR LEIVA²

¹DEPARTAMENTO DE MATEMÁTICA, INSTITUTO TECNOLÓGICO DE BUENOS AIRES, ARGENTINA. ²ESCUELA DE INGENIERÍA INDUSTRIAL, PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO, CHILE.

ccastrok@gmail.com

Datos censurados surgen en forma natural en ciencias de la salud y de la ingeniería, entre otras áreas. La censura ocurre cuando algunas unidades bajo análisis no han presentado el evento de interés antes de finalizado el estudio. En censura Tipo I, el tiempo de estudio es fijo y el número de unidades que fallan es aleatorio. En censura Tipo II, el tiempo del estudio es aleatorio y el número de unidades que fallan es fijo. Existe un tipo de censura llamada híbrida que mezcla los dos tipos mencionados. La censura híbrida se ha usado con éxito, por ejemplo, en modelos de riesgos competitivos. Los esquemas de censura híbridos se han extendido recientemente también a un tipo de censura llamado progresiva que ha sido discutida ampliamente por varios conocidos autores. Estos nuevos esquemas de censura han producido la apertura de un vasto campo para la investigación, ya que incluso para distribuciones muy conocidas, como la Weibull o la gamma, no se han obtenido todavía resultados de bondad de ajuste y de inferencia bajo censura híbrida. La censura progresiva es más alentadora en este aspecto, puesto que se han obtenido estimaciones de parámetros para algunas distribuciones conocidas, como la exponencial. El objetivo de este trabajo es discutir distintos métodos de censura, algunos resultados inferenciales obtenidos para esos métodos y los desafíos que se presentan al estimar parámetros distribucionales bajo estos métodos de censura. En este trabajo se introducen también algunos de resultados sobre bondad de ajuste y estimación bajo censura progresiva que se encuentran en estudio por los autores.

Palabras claves: distribuciones de pr; censuras tipo i y ti; censuras híbrida y p

Datos faltantes

EFECTO DE VALORES FALTANTES EN ESTUDIOS LONGITUDINALES EN ADULTOS MAYORES

FERNANDO MASSA

INSTITUTO DE ESTADÍSTICA, UNIVERSIDAD DE LA REPÚBLICA.

fmassa@iest.edu.uy

Comprender el proceso de deterioro cognitivo global de los adultos mayores es de gran relevancia para una mejor planificación no sólo a nivel económica/familiar, repercutiendo de forma directa en los individuos afectados, sino también a nivel del sistema de salud y seguridad social de un país. Puesto que el deterioro se manifiesta a lo largo del tiempo, para su estudio son implementados modelos longitudinales. Sin embargo estos modelos tienen como desventaja que a lo largo del período de seguimiento, por diversas causas, algunos individuos desertan del estudio. Este abandono no sólo reduce el tamaño muestral sino que genera importantes sesgos en los resultados de los análisis estadísticos. El objetivo de este trabajo fue modelar el deterioro cognitivo global de un conjunto de adultos mayores a lo largo del tiempo. Para llevar a cabo este objetivo se trabajó con datos provenientes del estudio Origins of Variance in the Old-old: Octagenarian Twins que contaba con una cohorte de 702 individuos entre 79 y 98 años de edad. Para medir el deterioro global se utilizó el Mini Mental State Examination (MMSE), evaluado a lo largo del tiempo en intervalos de dos años. Se implementaron modelos conjuntos puesto que estos parten de la base de que el tiempo de sobrevida de los sujetos y sus resultados de MMSE están relacionados. Se observó que los valores de MMSE son afectados tanto por la edad de los individuos al inicio del estudio como por el nivel educativo de los mismos. Incluir el análisis de sobrevida como parte del proceso de deterioro cognitivo corrige las inferencias. No existen trabajos previos en el Uruguay relacionados con la metodología desarrollada en este trabajo por lo que se considera que este trabajo representa una primera aproximación a la metodología a ser aplicada tanto para la planificación en el sistema de salud y seguridad social como para el sistema educativo e incluso para el contexto de trasplante de órganos.

Palabras claves: datos longitudinales; modelo conjunto; datos faltantes; deterioro cognitivo

Diseño de experimentos

ANÁLISIS POR DISEÑO EXPERIMENTAL DE LA REMOCIÓN DE CASEINATO DE SODIO POR EC-EF

*PAULA V SARMIENTO, MAURICIO MATTALIA, DIEGO G SEMPRINI y
MIGUEL A ROSA*

GIDAIQ - UNIVERSIDAD TECNOLÓGICA NACIONAL, FACULTAD REGIONAL VILLA MARÍA, VILLA MARÍA, CÓRDOBA.

paula.victoria.sarmiento@gmail.com

En este trabajo se aplicó un diseño de experimentos para obtener un modelo matemático representativo de la remoción de proteínas en efluentes mediante el proceso de electrocoagulación-electrofloculación (EC-EF). Además, se determinó un modelo que permitiese estimar el consumo de energía del proceso. Se aplicó el método de superficie de respuesta con un diseño de Box-Behnken el cual facilita la posterior optimización del modelo y presenta un menor número de experimentos respecto al diseño central compuesto. Las experiencias se realizaron empleando un efluente sintético, compuesto por caseinato de sodio (CS) y agua, el contenido de proteínas se determinó aplicando el método de Bradford. El reactor empleado consiste en una cuba de acrílico y una celda de 4 electrodos de aluminio con separación de 1 cm, área efectiva de 270 cm² y volumen útil de 1,7 L. La operación de la celda se realizó mediante una conexión bipolar, inversión de polaridad cada 5 minutos y agitación magnética. Se registró la intensidad de corriente que circulaba por la celda cada 1 minuto para calcular la energía consumida. El post-tratamiento consistió en centrifugar las muestras durante 10 minutos y determinar la absorbancia correspondiente al contenido de proteínas

remanente. Los factores seleccionados fueron concentración de CS, voltaje y tiempo con tres niveles. Se obtuvo un diseño con 17 corridas en un bloque. Para la remoción de proteínas se obtuvo un modelo cuadrático reducido donde los términos más significativos fueron el voltaje, el tiempo y el término cuadrático de la concentración. La falta de ajuste presentó un valor-p de 0.2759, lo cual indica que la misma no es significativa y el modelo ajusta correctamente los datos. El valor de R² fue de 93.87%, y el modelo desarrollado puede emplearse para uso predictivo dentro del espacio de trabajo analizado. Para obtener un modelo adecuado para la energía consumida, se debieron transformar los valores reales mediante una función raíz cuadrada. El mejor ajuste de los datos transformados se obtuvo con un modelo que incluye las interacciones entre dos factores, donde los más significativos fueron voltaje, tiempo, y la interacción de segundo orden entre los mismos. Este modelo arrojó un valor-p para la falta de ajuste de 0.8403 y un R² de 98.42%. El análisis de los gráficos de diagnóstico permitió confirmar su validez. Como próxima actividad se plantea realizar la optimización de estos modelos y verificar su eficacia de predicción.

Palabras claves: diseño de experimento; electrocoagulación; efluentes; caseinato de sodio

EXPERIMENTOS DE SCREENING BASADOS EN FRACCIONES NO REGULARES

MAITE SAN MARTIN, LUCIA HERNANDEZ y JOSE PAGURA

FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA - UNIVERSIDAD NACIONAL DE ROSARIO.

mai.sanmartin@gmail.com

En muchas investigaciones, y en particular en las primeras etapas de las mismas, el objetivo se centra en determinar cuáles factores, de un gran grupo, afectan a una variable de interés. Los diseños factoriales completos a dos niveles, conocidos como diseños 2^k , son planes experimentales que permiten estimar los efectos lineales asociados a k factores y todas las interacciones posibles entre dichos efectos. Sin embargo, su ejecución puede resultar muy costosa dado que, en general, requieren un gran número de pruebas. La alternativa tradicionalmente utilizada es la de llevar a cabo un diseño factorial fraccionario $2^{(k-p)}$, el cual consiste un subconjunto de los ensayos del diseño 2^k seleccionado apropiadamente. Estos diseños son construidos definiendo relaciones entre los efectos que conducen a estructuras de confusión simples (pares de efectos ortogonales o completamente confundidos). El número de tratamientos asociados a estos diseños es una potencia de 2, lo que resulta en cierta inflexibilidad en cuanto al número de pruebas a realizar. En los últimos años, los diseños factoriales fraccionarios no regulares han ganado popularidad en los experimentos de screening debido principalmente a su flexibilidad con respecto al número de ensayos que requieren y al número de niveles de los factores que pueden considerar. Estos diseños, a diferencia de los diseños $2^{(k-p)}$, pueden exhibir estructuras de confusión complejas donde los efectos pueden estar parcialmente confundidos introduciendo una complejidad adicional a su análisis. En este trabajo se presenta una clase particular de diseños factoriales fraccionarios no regulares llamados arreglos ortogonales y en particular a dos niveles y se enuncian algunos conceptos definidos en la literatura para estudiar sus propiedades. Además, se construyen diseños para distintos números de factores y de pruebas tanto mediante el enfoque tradicional como el de arreglos ortogonales con el objetivo de mostrar las ventajas y desventajas de cada clase.

Palabras claves: experimentos a dos n; diseños factoriales; arreglos ortogonales

UN ESTUDIO COMPARATIVO DE TRES CRITERIOS DE D-OPTIMALIDAD BAYESIANOS EN UN MODELO NO LINEAL

VÍCTOR IGNACIO LÓPEZ RÍOS

UNIVERSIDAD NACIONAL DE COLOMBIA SEDE MEDELLÍN.

vlopez@unal.edu.co

Los diseños de experimentos D-óptimos Bayesianos dan las condiciones experimentales donde un investigador debe conducir un experimento para estimar los parámetros de un modelo bajo estudio. Estos diseños resultan de la minimización de un criterio de optimalidad obtenido a partir de una función de utilidad definida apropiadamente. Bajo ciertos supuestos de normalidad e independencia en el término de error, dicho criterio se reduce a la optimización de un funcional de valor real de la matriz de información de Fisher. En la literatura de diseños óptimos existen distintas formulaciones para el criterio D-optimalidad Bayesiano. En este trabajo se comparan los diseños óptimos obtenidos bajo tres formulaciones distintas del criterio D-optimalidad, usando cuatro distribuciones a priori: uniforme discreta, uniforme continua, gamma y lognormal en un modelo exponencial de dos parámetros. Como estrategia metodológica se seleccionan los parámetros de las distribuciones consideradas de tal forma que la media y la varianza de éstas coincidan con la media y varianza de la distribución uniforme continua, ésta última justificada en la viabilidad que el investigador proporcione al menos los valores mínimo y máximo donde pueden variar los parámetros del modelo. Se obtienen los siguientes resultados: Se hallan todos los diseños óptimos para las distribuciones a priori consideradas usando diferentes escenarios para el soporte de los parámetros del modelo; adicionalmente, se obtienen las eficiencias de los diseños obtenidos y también se verifica que los diseños obtenidos son diseños óptimos usando el respectivo teorema de equivalencia. Se concluye que, para cada una de las distribuciones consideradas, los diseños obtenidos difieren considerablemente dependiendo de la versión del criterio de D-optimalidad usado. Se lograron obtener diseños de dos y tres puntos de soporte.

Palabras claves: distribuciones a pri; diseños d-óptimos ba; criterios de optimal; eficiencia; teorema de equivalen

Estadística espacial

ANÁLISIS ESPACIAL DE PARÁMETROS MAGNÉTICOS

MARCELA NATAL¹, MAURO CHAPARRO², LILA RICCI¹, MARCOS CHAPARRO³ y DÉBORA MARIÉ⁴

¹CENTRO MARPLATENSE DE INVESTIGACIONES MATEMÁTICAS (CEMIM-UNMDP), MAR DEL PLATA, ARGENTINA. ²CENTRO MARPLATENSE DE INVESTIGACIONES MATEMÁTICAS (CEMIM-UNMDP), MAR DEL PLATA, ARGENTINA. CONICET. ³CENTRO DE INVESTIGACIONES EN FÍSICA E INGENIERÍA DEL CENTRO DE LA PCIA. DE BUENOS AIRES. (CIFICEN, CONICET-UNCPBA), TANDIL, ARGENTINA. ⁴CENTRO DE INVESTIGACIONES EN FÍSICA E INGENIERÍA DEL CENTRO DE LA PCIA. DE BUENOS AIRES. (CIFICEN), TANDIL, ARGENTINA.

mnatal@mdp.edu.ar

La modelización de datos espaciales es de gran interés en muchas áreas del conocimiento, entre ellas en las ciencias ambientales. Particularmente el estudio de la contaminación del aire en ambientes urbanos a través de técnicas del magnetismo ambiental ha sido ampliamente estudiado durante la

última década. La ciudad de Tandil, provincia de Buenos Aires tiene como una de las principales actividades económicas la industria metalúrgica existiendo 7 de las mismas emplazadas dentro del área urbana que pueden ser fuentes de contaminación. El objetivo de esta contribución es construir modelos geoestadísticos para analizar la contaminación atmosférica por medio de un índice de contaminación magnética (ICM) y parámetros magnéticos. El conjunto de datos de la ciudad de Tandil consiste de 180 muestras de líquenes a los que, además de su ubicación geográfica, se le realizaron determinaciones magnéticas relativas a concentración, mineralogía y tamaño del material magnético en laboratorio. La descripción espacial de diferentes zonas de estudio permite determinar fuentes de contaminación y representar gráficamente las zonas más afectadas por los contaminantes. Se diseñó una estrategia para construir una grilla regular a partir de los datos obtenidos dado que los puntos muestrados pertenecen a un grillado irregular. Para la construcción de los modelos se analizó normalidad y se determinaron los índices de Geary y Morán para corroborar la existencia de autocorrelación espacial. Además se determinaron puntos atípicos espaciales. Se utilizaron las transformaciones gpot, dado que las variables consideradas son asimétricas. El semivariograma experimental se ajustó por los modelos: exponencial, gaussiano, esférico y wave. Para la selección del mejor modelo se utilizó el método clásico de validación cruzada “leave-one-out” y sus estadísticos de diagnóstico: el error medio (ME), el error cuadrático medio (MSE) y el cociente de la desviación cuadrática media (MSDR). La predicción espacial fue realizada por medio de un kriging ordinario. Se determinó una grilla regular con una distancia aproximada de 200 mts, se obtuvo normalidad luego de la transformación gpot con un parámetro de 1.38, se detectaron tres puntos espaciales atípicos. A partir de los modelos ajustados se construyeron los mapas de predicción. En las representaciones, en general, se observó que las zonas más contaminadas están cercanas a las metalurgias. Para los cálculos y gráficos se utilizó el paquete geoR del software libre R.

Palabras claves: medidas de asociación; modelo geoestadístico; transformación gpot; contaminación ambiental; grillado regular

COMPARACIÓN DE DIFERENTES METODOLOGÍAS PARA LA PREDICCIÓN DEL VALOR CONTADO DE UN INMUEBLE

LEONARDO MORENO, MARCO SCAVINO y JUAN JOSÉ GOYENECHE

UDELAR.

mrleo@iestra.edu.uy

El objetivo del trabajo es comparar la eficiencia predictiva de diferentes metodologías basadas en información espacio-temporal de un conjunto de inmuebles. Se tiene como insumo información espacial y temporal sobre un conjunto de inmuebles y ciertas variables intrínsecas a cada bien, por ejemplo, superficie, antigüedad, número de dormitorios. En tal sentido es conocido el valor contado (en dólares norteamericanos) de ciertos inmuebles de la ciudad de Montevideo (Uruguay) en diferentes fechas, entendiendo por valor contado a aquel que es asignado por el tasador, donde cada propiedad puede haber sido tasada en diferentes momentos, problema denominado en la literatura como ventas repetidas, [Bailey et al., 1963]. La información espacial, coordenadas de cada inmueble, permite la construcción de modelos autoregresivos espaciales donde el precio de un inmueble se encuentra correlacionado con el de sus vecinos, [Anselin, 1988]. La información temporal puede ser modelada mediante un modelo autoregresivo temporal, en este caso la dependencia entre una tasación y la siguiente disminuye en función al lapso de tiempo transcurrido entre una y otra, [Nagaraja et al., 2011]. Las variables hedónicas se modelan mediante un modelo de regresión lineal dinámico, donde los coeficientes de la regresión son función de la ubicación del inmueble, [Sun et al., 2014]. Es posible mediante la metodología de Stacking, [Breiman, 1996] encontrar un pro-

cedimiento que permita construir un único algoritmo que determine predicciones más precisas en términos del error cuadrático medio. En este caso, otras metodologías como los modelos mixtos, [McCulloch and Searle, 2001], y el modelo de Case-Shiller, [Shiller and Case, 1987], pueden ser incluídas en la agregación. Se evalúa la performance del modelo en datos simulados y en datos reales mediante validación cruzada. Es desafiante la búsqueda de un único modelo con una mejor performance que los ya desarrollados.

Palabras claves: método de stacking; modelos espacio-temp; precio hedónico; procesos autoregresi

IMPUTACIÓN DE DATOS FALTANTES DEL CENSO DE POBLACIÓN Y VIVIENDA UTILIZANDO TÉCNICAS DE ESTADÍSTICA ESPACIAL

MARÍA EUGENIA RIAÑO

SUE.

rianomiranda@gmail.com

En general, la calidad y cobertura de los Censos de Población y Vivienda del año 2011 realizado en Uruguay, fue calificada como positiva, cumpliendo con los estándares exigidos internacionalmente. Sin embargo, su implementación no estuvo exenta de inconvenientes. No se cuenta con información de determinados hogares cuyo domicilio fue relevado, y para algunos se cuenta con sólo información parcial relativa a la composición del hogar. La omisión censal se concentra en zonas socioeconómicamente vulnerables. Esto afectaría la construcción del mecanismo utilizado por el Ministerio de Desarrollo Social para seleccionar a la población beneficiaria de los programas de transferencia monetaria. Este mecanismo se basa en la Encuesta Continua de Hogares cuyo marco muestral es el del Censo, y refleja los problemas de omisión. El trabajo se desarrolla para la ciudad de Montevideo. El patrón espacial de la población objetivo y de la propia omisión hace necesaria una regionalización previa a la imputación, dado que la distribución espacial se muestra heterogénea en el mapa. La selección de los modelos a utilizar para la imputación es muy sensible a la escala del mapa, por lo que la definición de las regiones condiciona la selección del modelo final a utilizarse para realizar la imputación. Las regiones se construyen mediante el algoritmo de árboles oblicuos de decisión, implementado en el paquete SPODT de R. Se ajustan modelos autorregresivos espaciales en cada región (SAR) que son evaluados con métodos de validación cruzada. Las imputaciones se realizan como predicción de los modelos mejores evaluados. Los resultados muestran una mayor omisión de hogares socioeconómicamente vulnerables, lo que impacta directamente en la selección de población beneficiaria de programas de transferencia monetaria.

Palabras claves: imputación de datos; modelos sar autorreg; validación cruzada; árboles de clasifica

MEDIDAS TIPO R-CUADRADO PARA MODELOS ESPACIALES

LILA RICCI¹ y SILVIA OJEDA²

¹CENTRO MARPLATENSE DE INVESTIGACIONES MATEMÁTICAS, UNIVERSIDAD NACIONAL DE MAR DEL PLATA, ARGENTINA.

²FACULTAD DE MATEMÁTICA ASTRONOMÍA Y FÍSICA. UNIVERSIDAD NACIONAL DE CÓRDOBA.

lricci@mdp.edu.ar

En el proceso de ajuste de modelos espaciales a un conjunto de observaciones, resultará de gran

utilidad contar con medidas que permitan evaluar el ajuste global y ayuden en el proceso de selección de covariables. Para tener un criterio de admisibilidad de las medidas propuestas, Cameron y Windmeijer (1995) enunciaron las cinco propiedades que debe tener una medida de ajuste global para ser considerada “tipo R2”. Para Modelos Lineales Generalizados Zheng (2000) analiza cuatro coeficientes posibles; todos ellos se reducen al R2 clásico cuando los errores son normales. Es de particular interés la medida basada en remplazar las distancias euclídeas, que aparecen en el modelo clásico dentro de las sumas de cuadrados, por la deviance que es equivalente a la distancia de Kullback-Leiber. ***justificar*** Para el modelo normal multivariado, una medida de ajuste global está dada por el conocido estadístico lambda de Wilks, que es un cociente entre determinantes de las matrices de covarianza de dos hipótesis anidadas; varía entre 0 y 1 y es proporcional a la estadística F para la hipótesis de nulidad de todos los coeficientes. Nagelkerke (1991) amplió la definición del coeficiente de determinación a todo aquel que posea propiedades similares a las definidas por Cameron y Windmeijer. En este trabajo se resumen las diferentes medidas de ajuste global que han sido propuestas para modelos espaciales y se propone una medida que tiene las propiedades necesarias para ser considerada una medida de tipo R2, la misma está basada en la deviance, y tiene como antecedentes medidas similares propuestas para Modelos Lineales.

Palabras claves: estadística espacial; medidas de ajuste gl; modelos lineales

Inferencia estadística

AN EXTENSION OF HORVITZ-THOMPSON ESTIMATOR APPLICATION:

ALEJANDRO IZAGUIRRE¹ y GABRIEL MONTES ROJAS²

¹UNIVERSIDAD DE SAN ANDRÉS, BUENOS AIRES, ARGENTINA. ²INSTITUTO INTERDISCIPLINARIO DE ECONOMÍA POLÍTICA DE BUENOS AIRES AND CONICET, FACULTAD DE CIENCIAS ECONÓMICAS, UBA.

izaguirre.ale@gmail.com

Most networks investigated today are parts of a much larger networks. There are many papers in which the focus is on understanding the extent to which characteristics of a sampled network are reflective of the complete network. A central statistical question in those works is, how much the properties of the sampled network reflect those of the true network. Typical characteristics of interest include degree distribution, density, diameter, clustering coefficient, average path length, among others. In general, under many sampling schemes, these measures are biased when we use sampled networks as if they were the true. In this paper we focus our attention on the degree distribution, this measure is one of the most fundamental characteristic associated with a network and, it may be affected by sampling, sometimes dramatically. In literature there exist many different networks sampling designs which can be probabilistic or nonprobabilistic, here we restrict our attention on probabilistic samples, as in the rest of the works related with this topic. In the first part we present an extension of Horvitz-Thompson estimator(HT) to be used under more general context, basically with partial information on the variable of interest. Briefly, suppose you have a population N and your interest is in variable y_i , suppose you want to estimate the total of y_i , T_y , by a sample of N, under some assumptions you could use the HT estimator, but to do so you have to observe y_i in the sample. What we do in the first part of the work is to propose an extension of HT estimator to make it feasible under partial information on y_i . Based on the previous methodology, in a second part, we propose an estimator for degree distribution (and its variance) to be used under any probabilistic sampling designs, and then we apply it for two widely used probabilistic sampling designs known

as induced subgraph (nodes sampling) and incident subgraph (edges sampling). Finally we do a Monte Carlo simulation to assess some aspects of the theoretical results.

Palabras claves: horvitz-thompson est; networks; network sampling des; degree distribution; average degree

EL VALOR DE EDUCARSE EN ARGENTINA UN ANÁLISIS TEMPORAL

PAULA LORENA NAIDICZ, GABRIEL AMÓS FRIDRIJ y VÍCTOR FABIO LAZARTE

UNIVERSIDAD NACIONAL DE TUCUMÁN- FACULTAD DE CIENCIAS EXACTAS Y TECNOLOGÍA.

lorenafridrij@hotmail.com

El valor de educarse no abarca solamente los aspectos económicos y laborales de la vida de las personas, sino que nos forma para llevar una vida mejor en general, para uno mismo y para la sociedad. Una de las maneras de cuantificar la influencia del estudio en la calidad de vida de las personas es a través de un análisis del ingreso. Existen muchos trabajos en donde se evidencia que una mejor educación contribuye a tener mejores ingresos. El presente trabajo muestra que, en realidad, esta relación fue disminuyendo con el tiempo. Es decir si bien hay un efecto positivo entre la educación y el ingreso, éste es cada vez menor. Este trabajo presenta un análisis temporal con datos de la Encuesta Permanente de Hogares desde 2003 a la fecha, que analiza las relaciones existentes entre el nivel educativo e ingresos mensuales considerando la población económicamente activa. Los simples cocientes de los ingresos entre los diferentes niveles educativos, en cada trimestre, representan la cantidad de veces que el salario del nivel más bajo está contenido en el nivel más alto, por ejemplo si uno hace el cociente entre el ingreso mediano del secundario completo con primaria completa y el resultado es 3, quiere decir que el ingreso mediano del secundario completo es el triple que el del primaria completa. Si el cociente diera próximo a uno, significa que no hay diferencias entre completar uno u otro nivel. Este análisis permite ver que los ingresos son superiores para niveles educativos más altos en un instante de tiempo por aglomerado, pero el análisis de estos a través del tiempo, demuestra una tendencia decreciente significativa.

Palabras claves: nivel educativo; ingresos; brecha de ingresos; análisis temporal

UNA APLICACION ACTUARIAL DEL MODELO DE AZBEL

MARIA TERESA CASPARRI, MARIA ALEJANDRA METELLI y CAROLINA ARTUSO

FCE-UBA.

ametelli@gmail.com

Resumen En diversas circunstancias muchos profesionales, entre ellos especialmente los actuarios se enfrentan con la necesidad de conocer, no sólo cuál es la fuerza de mortalidad a la que está sometida una población determinada en un momento dado del tiempo, sino también conocer su evolución futura. Por dicho motivo, en el presente trabajo nos propusimos proyectar en el tiempo las tablas de mortalidad de Argentina para ambos sexos. Dado que la información disponible en nuestro país es escasa, fue necesario recurrir a fuentes externas. Desde principios del siglo XX, año tras año, muchos países han venido construyendo sus propias tablas de mortalidad. Haciendo uso

de dicha información, buscamos vincular la última tabla de mortalidad local disponible con alguna de las tablas norteamericanas que se han publicado desde el año 1933 al presente. El objetivo fue encontrar entre todas ellas y para cada sexo, una a la cual, la nuestra podría ser equivalente. Para ello, hemos recurrido al uso de tests estadísticos, y una vez completados los mismos hemos podido decidir por cuál podría ser reemplazada. Por otro lado, con el objetivo de analizar la mortalidad de la población de Estados Unidos, hemos utilizado al modelo de Azbel. Para cada una de las tablas mencionadas discriminadas por sexo, y mediante el método de mínimos cuadrados, hemos realizado una regresión lineal entre el logaritmo natural de la tasa anual de mortalidad y la edad. Una vez obtenida la estimación de los parámetros para cada uno de los años involucrados, y mediante la utilización de series de tiempo, éstos fueron proyectados de manera tal de conocer sus valores futuros. Completado dicho proceso, hemos podido construir las tablas de mortalidad venideras para dicho país. Por último, haciendo uso de dicha predicción, y suponiendo que el desarrollo temporal de ambas tablas es similar, logramos obtener las tablas de mortalidad proyectadas para Argentina.

Palabras clave: Proyecciones, Azbel, Mortalidad, Área temática: Inferencia estadística, aplicaciones actuariales.

Palabras claves: proyecciones; azbel; mortalidad

Métodos bayesianos

APLICACIÓN DEL ALGORITMO EM (EXPECTATION-MAXIMIZATION)

TULIA EVA SALCEDO PALACIOS y WILMER DARIO PINEDA RIOS

tuliasalcedo@usantotomas.edu.co

En estadística, el algoritmo de esperanza-maximización o EM es un método iterativo para encontrar las estimaciones por máxima verosimilitud o la probabilidad de un máximo a posteriori (MAP) de parámetros en modelos estadísticos, donde éste depende de variables latentes no observadas. Para el presente trabajo se ilustrará la utilización del algoritmo EM a través del paquete astsa para estimar la señal subyacente de la banca en Colombia, esto se hará a través de una recursión hacia atrás de un filtro de Kalman, que nos permite determinar los parámetros de un modelo dinámico lineal. Las series empleadas fueron la señal bancaria obtenida por (Saavedra et al. 2016), el IPI desestacionalizado y el IMACO, el sistema se describe mediante un modelo lineal y los ruidos de las mediciones se asumieron gaussianos para garantizar la convergencia a un máximo. Finalmente, empleando los hiperparámetros calculados mediante el algoritmo EM, se realizó el suavizado a fin de mostrar la tendencia de las series con un intervalo de confianza del 95%. El algoritmo EM es un proceso iterativo bastante práctico cuando se cuenta con variables no observadas, su simplicidad nos permitió en pocas líneas de código y con mucha rapidez calcular el conjunto de hiperparámetros que se buscaba hallar. Con estos datos se estimó la señal subyacente de la banca colombiana con la cual es posible llegar a conocer la interrelación entre las variables, ver los efectos que puede tener una sobre las otras variables analizadas, y generar pronósticos. Otros métodos empleados, como la inicialización difusa, suelen resultar incongruentes a la hora de justificar un estudio, ya que asumen varianzas infinitas al estado inicial con valores de magnitud observados finitos, con lo que el algoritmo EM no sólo ofrece una ventaja no sólo aplicativa sino también argumentativa.

Palabras claves: algoritmo em; filtro de kalman; señal bancaria

BAYESIAN ESTIMATION IN THE ADDITIVE HAZARDS MODEL

ENRIQUE E ÁLVAREZ¹ y MAXIMILIANO L RIDICK²

¹UNLP - UBA.IC - CONICET. ²UNLP - CONICET.

enriquealvarez@fibertel.com.ar

Suppose we have a sample of n individuals who may experience an terminal event over a window $[0, u]$. We denote by T_i the true, possibly latent, time to occurrence for the i -th individual. Because some individuals experience censoring at times C_i , their duration until the event is observed only when $C_i \geq T_i$. In classical Survival Analysis, it is of interest to study the soujourn times as related to observed individual covariates, which we assume time-independent and denote by Z_i . In the literature, models for survival data typically focus on the so-called hazard rate, which we assume takes the additive form $l(t, b) = l_0(t) + z^T b$ due to Aalen (1980), where $l_0(\cdot)$ is the baseline hazard function and b is a vector of unknown coefficients. Alternative approaches abound in the literature, the most common being Coxs (1972) proportional hazards model an the Accelerated failure time model. It is our goal in this study to propose a Bayesian method of estimation for the semiparametric Additive Hazards Model (AHM) under right-censoring. With this aim, we review the AHM and introduce the likelihood function, so that we comment on the challenges posed by estimation from the full likelihood. Thus we discuss an alternative approach based on a hybrid Bayesian method that exploits Lin and Yins (1994) estimating equation approach and simple tractable priors for the parameters. For the simplest case, we obtain the posterior distributions and we provide algorithms for the estimators. We illustrate our method with a simple dataset in the medical field. References
Aalen, O. O. (1980). A model for non-parametric regression analysis of counting processes. Lecture Notes on Mathematics Statistics and Probability. 2: 1-25.
Cox, D.R. (1972). Regression Models and Life-Tables (with discussion). Journal of the Royal Statistical Society. Series B. 34: 2, 187-220.
Lin, D. Y., Ying, Z. (1994). Semiparametric analysis of the additive risk model. Biometrika, Vol. 81, No. 1, pp. 61-71.

Palabras claves: bayesian inference; survival analysis; statistical inferenc; additive hazards mod

HIERARCHICAL BAYESIAN ANALYSIS OF ALLELE-SPECIFIC GENE EXPRESSION DATA

IGNACIO ÁLVAREZ-CASTRO¹ y JARAD NIEMI²

¹IESTA-UDELAR. ²IOWA STATE UNIVERSITY.

nachalca@gmail.com

Diploid organisms have two copies of each gene (alleles) that can be separately transcribed. The RNA abundance associated of any particular allele is known as allele-specific expression (ASE). When two alleles have sequences of polymorphisms in transcribed regions, ASE can be studied with RNA-seq read count data. Reads counts that can be unambiguously attributed to a specific allele are positively associated with that alleles expression. In plant breeding, hybrids are developed to take advantage of the genetic phenomenon known as heterosis or hybrid vigor. Heterosis occurs when hybrid offspring possess superior levels of one or more traits relative to their inbred parents. ASE is relevant for the study of this phenomenon at the molecular level. One possible reason for the occurrence of heterosis are genes where two distinct alleles at a heterozygous locus are differentially expressed. ASE has some characteristics different from the regular RNA-seq expression: ASE cannot be assessed for every gene, measures of ASE can be biased towards one of the alleles (reference

allele), and ASE provides two measures of expression for a single gene for each biological samples with leads to additional complications for single-gene models. We present statistical methods for modeling ASE and detecting genes with differential allele expression. We propose a hierarchical overdispersed Poisson model to deal with ASE counts. The model accommodates gene-specific overdispersion, it has an internal measure of the reference allele bias, and uses random effects to model the gene-specific regression parameters. Fully Bayesian inference is obtained using the fbseq package that implements a parallel strategy to make the computational times reasonable. Simulation and real data analysis suggest the proposed model is a practical and powerful tool for the study of differential ASE.

Palabras claves: hierarchical models; allele-specific expr; full bayesian; shrinkage priors; multiple comparison; reference genome bia

INTEGRANDO LA ESTIMACIÓN BAYESIANA A LOS MODELOS DE ECUACIONES ESTRUCTURALES (SEM)

ANDREA DE LOS ÁNGELES CÉSPEDES SOLÍS

UNIVERSIDAD DE COSTA RICA. SAN JOSÉ – COSTA RICA..

ancelis.07@gmail.com

El enfoque bayesiano es considerado una alternativa atractiva para la estimación de modelos que tiene información previa sobre un fenómeno de interés, así como un tamaño de muestra pequeño o problemas de convergencia en las estimaciones: recientemente las poderosas herramientas computacionales relacionadas con la simulación le han dado una gran aplicabilidad a esta técnica. Por otra parte los modelos de ecuaciones estructurales (SEM), conforman otra opción de análisis estadístico, su principal particularidad es generar una explicación que apoya o refuta un modelo definido bajo un marco teórico sólido. A pesar de que éstas técnicas son útiles ante determinadas situaciones, su aplicación de manera integrada es escasa en la literatura, por lo anterior, en el presente trabajo realiza la estimación de un modelo de ecuaciones estructurales (SEM) bajo el enfoque bayesiano, a partir de las especificaciones dadas por Sik-Yum Lee (2007) en “Structural Equation Modeling: A Bayesian Approach”. La técnica es aplicada a un modelo previamente estimado con el enfoque frecuentista, denominado la violencia contra las mujeres por parte de sus esposos o compañeros: Un modelo estructural desde el enfoque de género y la psicología evolucionaria, inicialmente propuesto propuesto por la Dra. Eiliana Montero Rojas de la Universidad de Costa Rica y el Dr. Aurelio José Figueredo de la Universidad de Arizona, USA, en el 2005, y replicado por la Dra. Eiliana Montero Rojas en conjunto con Br. Mauricio Campos Fernández, Br. Rebeca Sura Fonseca y Br. Andrea Céspedes Solís de la Universidad de Costa Rica en el 2016. Las estimaciones obtenidas en ambos casos son comparadas de tal manera que se dan a conocer los posibles beneficios al utilizar la estadística bayesiana en relación con la estimación convencional. Para la estimación del modelo SEM bayesiano, son asignadas a las distribuciones priori varianzas pequeñas, específicamente la distribución Gamma con parámetros 4 y 9 y la Wishart. Además, se elaboran dos cadenas considerando 4 000 iteraciones de burn-in y 6 000 más para la estimación de los parámetros. Es importante mencionar que parte del conocimiento previo del modelo considerado para la definición de los hiperparámetros, es tomado de los resultados de la estimación frecuentista con la que ya se contaba. Para la estimación del modelo se utilizan los programas LISREL 8.80 y OpenBUGS 3.2.3. Finalmente, en la estimación de algunos de los parámetros, el resultado es similar en ambos enfoques.

Palabras claves: ecuaciones estructur; cadenas de markov ví; estadística bayesian; distribuciones a pri; hiperparámetros

MODELO DINA APLICADO A LA EVALUACIÓN DE MATEMÁTICA EN ESTUDIANTES DE SEGUNDO GRADO DE SECUNDARIA

YURIKO KIRILOVNA SOSA PAREDES y LUIS VALDIVIESO SERRANO

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ.

yuriko.sosa.p@gmail.com

Los modelos de diagnóstico cognitivo (MDC) tienen como finalidad describir o diagnosticar el comportamiento de los evaluados por medio de clases o perfiles latentes, de tal manera que se obtenga información más específica acerca de las fortalezas y debilidades de ellos. Uno de los modelos más populares de esta gran familia es el llamado modelo DINA, el cual tuvo su primera aparición en Haertel (1989) enfocado principalmente en el campo educacional. Este modelo considera solo respuestas observadas dicotómicas de parte de los individuos y tiene como restricción principal que ellos deben dominar necesariamente todas las habilidades requeridas por cada ítem; aquellas que se resumen en una matriz llamada Q. Asimismo, el modelo estima parámetros para los ítems, los cuales son denominados de ruido: Adivinación y Desliz. En este trabajo se desarrolla teóricamente el modelo expuesto; es decir, sus fundamentos y principales propiedades desde el enfoque bayesiano. Específicamente, las estimaciones se realizan mediante el Muestreador de Gibbs. Se presenta una aplicación enfocada en educación, donde se analiza una muestra de 3040 alumnos del 2do grado de secundaria, evaluados en una prueba de 48 ítems de la competencia matemática realizada por la Oficina de Medición de la Calidad de los Aprendizajes (UMC) en el 2015. A esta prueba se le aplica el modelo de Rasch y el modelo DINA bajo el enfoque bayesiano, con el fin de estudiar la correspondencia entre indicadores de ambos modelos, tanto para los parámetros de los alumnos (habilidad y perfiles latentes) como de los ítems (dificultad y parámetros de ruido).

Palabras claves: modelo dina; mdc; muestreador de gibbs; modelo rasch; variables latentes; educación

PRONÓSTICO DEL PRECIO DE ACCIONES EMPLEANDO MODELOS OCULTOS DE MARKOV

LUIS DAMIANO

UNR.

damiano.luis@gmail.com

El pronóstico fuera de la muestra es posiblemente el problema estadístico de mayor interés en el ámbito de las series de tiempo financieras. Se emplea la inferencia Bayesiana para estimar los parámetros de un Input-Output Hidden Markov Model (IOHMM) aplicado al pronóstico de precios de acciones argentinas. IOHMM es una arquitectura que relaciona una secuencia de entrada, o señal de control, con una secuencia de salida. Se trata de una variante con supervisión de los Modelos Ocultos de Markov (HMM), uno de los métodos de aprendizaje automático no supervisado (Unsupervised Machine Learning) más empleados para series de tiempo. Se asume la existencia de estados latentes para los precios, los cuales se modelan con una mezcla de gaussianas. Estimados los parámetros, se aplica una medida de distancia sobre la log-verosimilitud estimada de cada elemento dentro de la muestra a fin de identificar patrones en la secuencia histórica y producir un pronóstico replicando este comportamiento fuera de la muestra. La estimación bayesiana se realiza a través del método de cadenas de Markov Monte Carlo (MCMC). Se implementa el algoritmo de simulación de muestras en Stan, el lenguaje de programación probabilística más moderno

e influyente. Los estados latentes dan lugar a una función de densidad a posteriori con marcada multimodalidad, agravando las frecuentes dificultades que las mezclas de distribuciones presentan por la falta de identificabilidad. Se mejora la convergencia incorporando creencias a priori en la estructura del modelo estableciendo una relación jerárquica en los parámetros de locación de los componentes que conforman cada estado latente. Se divide la muestra en los conjuntos de datos de entrenamiento y prueba, se estiman los parámetros del modelo y se realizan las predicciones fuera de la muestra empleando una ventana rodante de datos (walking-forward). Se obtienen medidas de error de pronóstico satisfactorias. Se observa con particular interés que los estados latentes resultan persistentes, una característica valiosa en el ámbito de las finanzas que modelos econométricos como los cambios de regímenes con cadenas de no logran replicar. Se destaca especialmente que la especificación es muy flexible, con gran potencial para adaptarse a otros requerimientos típicos en el análisis de series de tiempo financieras. Se insta a futuros trabajos a experimentar con otras variables de entrada como ser el volumen negociado, la volatilidad o indicadores de tendencia. Nota al evaluador: Para más información, <https://goo.gl/Sr2x1c>. Aplicación similar que desarrollé para Google Summer of Code.

Palabras claves: input-output hidden; series de tiempo; finanzas; pronósticos para el; métodos bayesianos; markov-chain monte c

Métodos multivariados

ANÁLISIS DE DATOS MULTIVARIADOS CON COORDENADAS PARALELAS: UNA APLICACIÓN AL ESTUDIO DE CAMBIO CLIMÁTICO MEDIANTE EL ANÁLISIS DE LA TEMPERATURA

CARLOMAGNO ARAYA ALPÍZAR

UNIVERSIDAD DE COSTA RICA.

carlomagnocr@gmail.com

En esta ponencia, se estudia las Coordenadas Paralelas (Coords), que son un sistema de visualización que permite representar n-dimensiones en un sistema bidimensional. En este sistema, cada eje vertical (ordenada) representa un atributo (dimensión) que puede ser continuo o categórico. Cada uno de los ejes verticales de un sistema de Coords puede tener su propia escala o definirse todos con una sola escala. La primera forma nos permite la visualización de hiper-superficies y el análisis del comportamiento del conjunto de datos, con la segunda podemos hacer un análisis de las relaciones entre las variables. En general, las Coords son una técnica de visualización donde las dimensiones son simbolizadas como una serie de ejes paralelos, con la misma separación entre ellos (equidistantes) y donde los valores son representados. Cada eje representa una coordenada en la dimensión correspondiente. Uniendo con líneas los ejes, podemos simbolizar los puntos en n-dimensiones. Asimismo, un punto en un espacio n-dimensional es transformado en una línea poligonal a través de n ejes paralelos como n-1 segmentos de línea. La exploración visual de datos multidimensionales es de gran interés en estadística y en la visualización de información. Para tal efecto, se aplican las Coords al estudio del cambio de la temperatura en la comunidad de San Ramón (Costa Rica), usando los datos de dos estaciones meteorológicas. El estudio se realizará para temperaturas medias diarias de cada estación. En conclusión, las Coordenadas Paralelas son un método multivariado para visualizar en un plano n-dimensional los cambios de las temperaturas por mes y

año entre 2008-2017. Como resultado hemos obtenido a través del Modelo Lineal General que la temperatura promedio por mes y año ha variado significativamente, mostrando patrones que son un reflejo del cambio climático.

Palabras claves: coordenadas paralela; visualización multid; cambio climático

ELABORACIÓN DE PERFILES EPIDEMIOLÓGICOS EN ESTUDIOS SANITARIOS MEDIANTE TÉCNICAS DE CLUSTERING DIFUSO Y ANÁLISIS DE REDES

RAMON ÁLVAREZ y FERNANDO MASSA

INSTITUTO DE ESTADÍSTICA - DMMCC - FCEA.

ramon@iestra.edu.uy

En los estudios epidemiológicos es práctica habitual trabajar con variables binarias que reflejan la presencia de determinadas enfermedades, las que a su vez se asocian con otro conjunto otras enfermedades, denominadas comorbilidades, medidas también a través de variables binarias y que en general se asumen como factores de riesgo de las primeras. En el ámbito de los estudios epidemiológicos existen situaciones donde se manejan enfermedades no transmisibles (ENT), en particular en salud bucal, donde ambos tipos de variables pueden ser intercambiables en cuanto a quien hace el rol de factor de riesgo. Teniendo en cuenta esta situación se proponen comparar dos análisis para la determinación de tipologías de encuestados en base a los atributos binarios, de forma de obtener perfiles epidemiológicos bien diferenciados. Los datos utilizados corresponden a un estudio en personas que demandan atención en la Facultad de Odontología-UDELAR durante el período 2015-2016. Por un lado se emplea un método mixto que combina clustering basado en medidas de entropía con una partición difusa estimada con el algoritmo c-modes y por otro lado, a través del análisis de redes (SNA), a partir de las mismas variables se construye la matriz de adyacencias sobre la que se aplican una batería de métricas (closeness, betweenness, modularity, clustering) sobre los nodos y enlaces, que permite detectar comunidades. Las comunidades detectadas mediante el SNA a través del uso de diferentes algoritmos de búsqueda, como el de fast greedy o de random walk o de particionado espectral, se cruzan con las tipologías creadas mediante el método mixto de clustering.

Palabras claves: análisis de redes; clustering; entropía; factores de riesgo; variables binarias

MÉTODOS DE ANÁLISIS MULTIVARIADO A TRES VÍAS PARA IDENTIFICACIÓN DE ASOCIACIONES ENTRE CARACTERES FENOTÍPICOS Y MARCADORES MOLECULARES

SERGIO BRAMARDI¹, SABRINA COSTA TÁRTARA² y RAMIRO CURTI³

¹DEPARTAMENTO DE ESTADÍSTICA, UNIVERSIDAD NACIONAL DEL COMAHUE. ²DEPARTAMENTO DE TECNOLOGÍA, UNIVERSIDAD NACIONAL DE LUJÁN. ³CÁTEDRA DE ESTADÍSTICA Y DISEÑO EXPERIMENTAL, FAC. DE CS. NATURALES, UNIVERSIDAD NACIONAL DE SALTA (SEDE METÁN).

sbramardi@gmail.com

En la conservación de recursos fitogenéticos, la evaluación del material vegetal es una etapa crítica para su aprovechamiento, la que generalmente culmina con una descripción y clasificación de las poblaciones bajo estudio. Es común que esta caracterización se realice en función de atributos morfológicos-agronómicos y marcadores moleculares. Utilizar esta información para avanzar en el

establecimiento de relaciones entre caracteres fenotípicos y los marcadores moleculares puede permitir una incipiente detección de loci asociados a estos caracteres con potencial uso en posteriores procesos de mejora. Estos análisis involucran estructuras de datos a tres vías, más precisamente, conjunto de datos múltiples, por cuanto un mismo conjunto de individuos son descriptos por distintos tipos de variables: cuantitativas, cualitativas y moleculares. Existen diferentes alternativas de estudio para estas situaciones como el Análisis de Procrustes Generalizados (APG), Análisis Factorial Múltiple (AFM) y Análisis Factorial de Correspondencias aplicado sobre los datos discretizados según Escofier. En estos tres métodos, con sustentos teóricos totalmente diferenciados, se presenta una etapa de análisis de las interrelaciones de las tres vías, para luego, construir una estructura de consenso o compromiso de los individuos caracterizados a través de los distintos tipos de variables. Quizás para el objetivo de establecer relaciones entre individuos en base a variables mixtas la técnica más difundida sea el APG, pero ésta no permite una representación de los caracteres como si lo hacen las otras dos. En el presente trabajo, a partir de datos correspondientes a la caracterización de 25 poblaciones de quinoa de la región del Noroeste de Argentina a las que se les observaron 23 variables cuantitativas, 11 cualitativas y 22 loci microsatélites (SSR) que originaron 313 alternativas alélicas, se compararon las configuraciones de individuos obtenidos por las tres técnicas y se establecieron posibles relaciones entre caracteres fenotípicos y marcadores en las configuraciones de variables obtenidas del AFM y el AFC. Estas relaciones fueron verificadas a través de Pruebas chi-cuadrado y Análisis de la Variancia.

Palabras claves: quinoa; datos de conjuntos m; análisis de procrust; análisis factorial m; discretización de es; microsatélites

PROPIEDAD DE MUESTREO DE ÍNDICES DE SIMILITUD BASADOS EN DATOS BINARIOS

PATRICIA S. TORRES¹, MARTA B. QUAGLINO² y SERGIO CAMIZ³

¹CÁTEDRA DE ECOLOGÍA, FACULTAD DE CIENCIAS AGRARIAS (UNR). ²ESCUELA DE ESTADÍSTICA, FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA (UNR). ³UNIVERSIDAD LA SAPIENZA, ROMA (ITALIA).

patrizia662@gmail.com

Los índices de similitud basados en datos binarios han sido ampliamente usados en Ecología de comunidades, ya sea, para agrupar las unidades experimentales como las especies de una comunidad biótica. Índices similares o idénticos fueron aplicados subsecuentemente en otras disciplinas biológicas. Como resultado del esfuerzo por cuantificar la asociación o similitud en varios campos de la Biología, han aparecido una gran cantidad de medidas. Se encuentran actualmente en la literatura más de 60 índices de similitud, aunque algunos son simplemente funciones de otros. Existen varias propiedades que deben cumplir los índices de similitud para ser considerados “buenos índices”. Entre ellas se encuentra una que se refiere a la propiedad de muestreo de los mismos, por lo tanto el objetivo de este trabajo es evaluar la propiedad de muestreo de 61 índices de similitud encontrados en la literatura, a partir de procedimientos de simulación. Para probar esta propiedad se siguió un esquema similar al utilizado por Goodall en 1973, se consideraron 5 escenarios con distintas probabilidades de presencia para las especies. Los 5 escenarios se corresponden con valores conocidos y constantes para las frecuencias de celdas: a, b, c y d. Se tomaron muestras con repetición de cada una de las poblaciones, siguiendo el esquema de muestreo de Goodall (1973): 100 muestras de tamaño 10, 100 de tamaño 20, 100 de tamaño 40 y 100 de tamaño 80, se calcularon los índices, se promediaron y se testaron las diferencias entre los valores poblacionales y los muestrales. En la mitad de los índices se encontraron más de un 50% de resultados viciados mientras que la otra mitad resultaron en general poco viciados. El mejor comportamiento fue para el índice de Dennis y el peor para el índice de Pierce.

Palabras claves: índices; similitud; muestreo; binarios; aleatorización

Métodos no paramétricos

PROJECTION PURSUIT CLASSIFICATION RANDOM FORESTS

NATALIA DA SILVA¹, DIANNE COOK² y LEE EUN-KYUNG³

¹IESTA, UDELAR. ²MONASH UNIVERSITY. ³EWHA WOMANS UNIVERSITY.

ndasilva@iastate.edu

A random forest is an ensemble learning method built on bagged trees with random predictor selection. These features provide improved classification models because they produce information about the variable importance, predictive error, and proximity of observations. A new ensemble learning method for classification problems called projection pursuit random forest (PPF) will be presented. PPF uses the PPtree algorithm introduced in Lee et al.(2013). In PPF, trees are constructed by splitting on linear combinations of randomly chosen variables. Projection pursuit is used to choose a projection of the variables that best separate the classes. Utilizing linear combinations of variables to separate classes takes the correlation between variables into account which allows PPF to outperform the traditional random forest when separations between groups occur in combinations of variables. Previous work using oblique trees in the forest construction have shown positive results in terms of performance but are only for two-class problems. The method presented here can be used in multi-class problems and is implemented into an R package, PPforest, which is available at <https://github.com/natydasilva/PPforest>. Additionally, a forest classifier is an example of an ensemble since it is produced by bagging multiple trees. The process of bagging and combining results from multiple trees produces numerous diagnostics which, with interactive graphics, can provide a lot of insight into class structure in high dimensions. Various aspects will be explored in this presentation, to assess model complexity, individual model contributions, variable importance and dimension reduction, and uncertainty in prediction associated with individual observations. The ideas are applied to the random forest algorithm and projection pursuit forest, but could be more broadly applied to other bagged ensembles. Interactive graphics are built in R using the ggplot2, plotly, and shiny packages.

Palabras claves: classification; projection pursuit; ensemble method; statistical visualiz; interactive graphics; shiny app

REPRESENTAÇÃO DA ESTRUTURA DO PENSAMENTO COLETIVO SOBRE AS ENCHENTES DO RIO DOCE: CONECTANDO INDIVÍDUOS AFINS ATRAVÉS DA TEORIA DOS GRAFOS

WESLEY PEREIRA¹, GILVAN GUEDES¹, DENISE DUARTE² y RODRIGO RIBEIRO³

¹UNIVERSIDADE FEDERAL DE MINAS GERAIS. ²UNIVERSIDADE FEDERAL DE MINAS GERAIS. ³INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA.

denise@est.ufmg.br

Introdução O município brasileiro de Governador Valadares é historicamente impactado pelas inundações (enchentes) do Rio Doce. As enchentes são fenômenos naturais que normalmente são sinônimos de prosperidade, uma vez que estas contribuem para a fertilidade das terras próximas ao rio. Porém,

as enchentes em áreas urbanas podem se transformar em eventos catastróficos quando combinadas com intervenções no meio ambiente capazes de catalisá-las. Segundo Moscovici, 1961, idealizador da Teoria das Representações Sociais, dependendo dos significados sobre o objeto enchentes difundidos na sociedade, as pessoas podem não considerar as enchentes como um perigo a elas mesmas. Representação da rede de indivíduos através da Teoria dos Grafos Nesta análise, cada indivíduo contribui com até cinco evocações para a estrutura coletiva de significados. Uma técnica de representação dessa estrutura é a representação da rede através de grafos. Na Teoria dos Grafos, um grafo $G = (V, E)$ é definido como um conjunto de objetos de duas naturezas: o vértice (V), a unidade fundamental sob a qual o grafo é construído e as arestas (E), que representam a ligação entre pares de vértices. A essas ligações podem ser atribuídas muitas características, como direção, distância e peso. Representar a rede de indivíduos por intermédio de um grafo possibilita evidenciar indivíduos “populares, os quais possuem um grande número de conexões, resumindo as ideias mais difundidas em sua comunidade. Outra propriedade desta abordagem consiste na possibilidade de encontrar a difusão espacial do pensamento coletivo, sob a condição de se conhecer o georreferenciamento dos indivíduos. Dessa forma, procura-se delimitar a área de propagação do pensamento, verificando se existe nele algum indício de influência geoespacial. Além disso, pode-se realizar a validação das partições encontradas, bem como a verificação sobre indícios de processos latentes responsáveis pelo regimento comportamental da rede. Conclusão O estudo realizado leva à conclusão de que existem diferenças evidentes na perspectiva que as pessoas têm ao significar as enchentes. Exemplificando, existem alguns subgrupos da sociedade que buscam identificar as causas das enchentes enquanto outros subgrupos se preocupam com suas consequências. Alguns subgrupos expressam os sentimentos ruins que estes eventos geram à sociedade, enquanto outros se concentram nos prejuízos resultantes. A união das perspectivas de cada um desses subgrupos formam o pensamento coletivo da sociedade: as enchentes do Rio Doce são um evento natural que, pela uma série de fatores, são significadas majoritariamente pelas pessoas com um evento prejudicial à população e ao município de Governador Valadares.

Palabras claves: representações socia; enchentes rio doce g; pensamento coletivo; teoria de grafos; comunidades em grafo

Minería de datos

EL ARTE DE LOS NÚMEROS EN LA TRIBUTACIÓN

ANA OÑA

UNIVERSIDAD CATÓLICA DE CHILE.

alona@uc.cl

La Ley de Benford, conocida también como la ley del primer dígito a lo largo del tiempo ha sido empleada para descubrir anomalías en la información en datos numéricos como por ejemplo declaraciones de impuestos. Su aplicación permite detectar datos erróneos que no siguen un patrón teórico, de acuerdo a esta ley la distribución de los números del 1 al 9 como primer dígito de una cifra es asimétrica. La frecuencia con la que inician las cifras con el número 1 es aproximadamente el 30% mientras que iniciar con el número 2 ya es del 17% y así va descendiendo. Esta Ley se ha empleado para detectar el fraude fiscal y aunque en concepto parece intuitiva da luces de posibles actos de evasión y mala declaración. El boom tecnológico, la globalización y las tendencias de internacionalización hacen que cada vez se tenga más información de las empresas tanto de forma directa como por información de terceros. Al mismo tiempo, esto provoca que con tanta información los

procedimientos tradicionales de control de las Administraciones Tributarias se vuelvan obsoletos y poco oportunos. Contar con herramientas que permitan que este control sea oportuno y eficiente se vuelve imprescindible. El realizar controles en un periodo de tiempo más cercano incrementa la percepción de riesgo y disminuye la opción de manipular los datos. La detección del fraude fiscal debe realizarse desde la declaración, de manera inmediata, no costosa y de fácil implantación. La presente investigación busca corroborar la distribución del primer dígito con las declaraciones de impuesto a la renta de las sociedades en Ecuador. A su vez, determinar esquemas para utilizarla como herramienta para detectar anomalías en las declaraciones de impuestos.

Palabras claves: aplicación ley de be; fraude fiscal; anomalías datos

INCORPORACIÓN DE LOS PROCESOS KDD (KNOWLEDGE DISCOVERY IN DATABASES) EN LA CARRERA DE MEDICINA VETERINARIA

**GABRIELA PILAR CABRERA, ANDREA PEÑA MALAVERA, LAURA
ANGÉLICA MARCO y ELINA DEL VALLE FERNÁNDEZ**

UNIVERSIDAD NACIONAL DE VILLA MARÍA.

gabriela.pilar.cabrera@gmail.com

En el marco de la asignatura “Metodología de la Investigación” de la carrera de Medicina Veterinaria de la Universidad Nacional de Villa María (UNVM), se desarrolló una experiencia didáctica que permitió introducir a los estudiantes en la aplicación del proceso de KDD (Knowledge Discovery in Data Bases); a la vez que vivenciar los conceptos y procedimientos ejes de dicha asignatura. El problema de estudio específico, se contextualizó en la implementación de los procesos KDD para la caracterización de calidad de leche bovina cruda en aspectos Higiénico-Sanitario, Composicional y Adulteración y sus variaciones estacionales de cinco empresas lácteas del departamento Río II, Provincia de Córdoba en 2014. Con los estudiantes, se implementaron las etapas de limpieza e integración, obtención del Data Warehouse, evaluación y presentación de patrones. En particular, el pre-procesamiento de los datos provistos por los usuarios (Laboratorio de Diagnóstico Villa María), con la asistencia de expertos en Calidad de Leche Bovina, posibilitó comprender el dominio del problema de estudio (entendimiento del negocio). En tanto, la evaluación de la calidad de los datos generados y la identificación de patrones en los mismos, se realizó mediante un análisis exploratorio con el soporte del software estadístico InfoStat. Este entorno de trabajo, permitió a los estudiantes re-significar y consolidar conocimientos y herramientas abordados en Bioestadística y a su vez re-conocerla como tecnología de la investigación científica. En la implicación de los estudiantes en un problema real de investigación, resultó significativa la asignatura Metodología de la Investigación. Esta experiencia didáctica, promueve la inclusión de contenidos introductorios a los procesos KDD, en los programas de estudio de las asignaturas Informática y Bioestadística; de manera conjunta e integrada

Palabras claves: minería de datos; pensamiento estadístico; carrera de grado

Modelos de regresión

ALTERNATIVA PARA LA ASIGNACIÓN DE VALORES DE REFERENCIA EN LA DIMENSIÓN SALUD REPRODUCTIVA DEL ÍNDICE DE DESIGUALDAD DE GÉNERO

ALDO J VIOLLAZ, JORGELINA MENA y VIVIANA LENCINA

INSTITUTO DE INVESTIGACIONES ESTADÍSTICAS, FACULTAD DE CIENCIAS ECONÓMICAS, UNT.

jorgelinamena@yahoo.com.ar

En la agenda de desarrollo sostenibles, acordados por los países miembros de la ONU en septiembre de 2015, el lograr la igualdad entre los géneros y empoderar a todas las mujeres y las niñas sigue siendo un objetivo a alcanzar. Los índices multidimensionales han demostrado ser instrumentos valiosos para captar desigualdades, colaboran generando atención, estimulan el debate y análisis de políticas desarrolladas en este sentido además ayudan a monitorear los progresos logrados. Una de las medidas desarrolladas y publicadas por Human Development Report desde 2010, es el Índice de Desigualdad de Género (GII) compuesto por tres dimensiones: Salud Reproductiva, Empoderamiento y Participación del mercado laboral. El GII ha tenido grandes repercusiones tanto positivas como negativas, puesto que incorporó nuevas dimensiones y conceptos omitidos en otros índices, pero también fue criticado por la complejidad de su definición y la incorporación en su construcción tanto de indicadores que se calculan para ambos géneros, como otros que sólo están definidos para el sexo femenino, específicamente, los utilizados en la dimensión Salud Reproductiva. Puesto que en esta dimensión los indicadores incorporados no son medidos para el género masculino se le asigna un valor ideal, de manera independiente a la situación general de la región en donde habitan, característica que penaliza a regiones económicamente menos favorecidas, dado que los indicadores utilizados en la contraparte femenina están asociados al bienestar económico de la región donde se aplica. En este trabajo se realiza una propuesta de modificación en la definición de la dimensión Salud Reproductiva. Haciendo uso de regresión de cuantiles se construye una frontera inferior de posibilidades para cada indicador incorporado en esta dimensión y este valor es el asignado a los hombres en la construcción del índice. Se analiza el comportamiento de esta propuesta metodológica en las provincias argentinas.

Palabras claves: desigualdad de género; salud reproductiva; regresión de cuantil

EFECTOS DE LAS TRANSFERENCIAS CONDICIONADAS DE INGRESO SOBRE LA PARTICIPACIÓN LABORAL DE LOS ADULTOS: EL CASO DE LA AUH EN ARGENTINA

MARIANA HEREDIA y GABRIEL WEIDMANN

FACULTAD DE CIENCIAS ECONÓMICAS - UNER.

meriheredia@hotmail.com

En el año 2009 Argentina puso en marcha el programa Asignación Universal por hijo (AUH) para Protección Social, el cual universaliza las asignaciones familiares a los segmentos de la población más vulnerables no amparados por el Régimen de Asignaciones Familiares. Los objetivos de este

programa son dos: incrementar el bienestar del sector más pobre de la sociedad y fomentar la acumulación de capital humano en los niños. En general, existe una vasta evidencia en la literatura que las Transferencias Condicionadas de Ingreso son exitosas en América Latina en cuanto a la reducción de la pobreza y producir mejoras en la distribución de ingreso. Sin embargo, existen otros aspectos que motivan debates, principalmente, el impacto de este tipo de políticas sobre la actividad laboral de los adultos en los hogares destinatarios. Tanto el modelo estándar de oferta laboral como el modelo de oferta laboral familiar, predicen que las transferencias de ingresos no laborales, producen desincentivos al trabajo dado el efecto ingreso de las transferencias. En nuestro país, la evidencia en este campo es aún escasa y en diferentes direcciones. Para investigar estos posibles efectos no deseados, se plantea evaluar los efectos de la asignación universal por hijo en la participación laboral de los adultos, la tasa desempleo y tasa de actividad, en el corto y mediano plazo, con el fin de realizar un aporte al diseño de una política pública más efectiva de protección social. La metodología aplicada es el estimador de diferencias en diferencias, combinado con un modelo logit ya que en los diferentes modelos, la respuesta es discreta. Se utilizan los microdatos de la encuesta anual que realiza el Observatorio Social de la Universidad Nacional del Litoral en la ciudad de Santa Fe en el periodo 2009-2012. Para ello, se comparan dos grupos de adultos, uno que declara ser beneficiario de la AUH en la encuesta, con otro de características muy similares pero que no recibe el beneficio. Los resultados sugieren que ni en el corto ni mediano plazo, la AUH tiene efectos no deseados en el mercado laboral.

Palabras claves: participación de la; transferencias de in; evaluación de políti; estimador diferen- cia

ESTIMACIÓN ROBUSTA EN MODELOS DE REGRESIÓN CON ERRORES EN LAS VARIABLES FUNCIONALES

PATRICIA GIMÉNEZ y LUCAS GUARRACINO

DEPTO DE MATEMÁTICA - FCEYN. UNIVERSIDAD NACIONAL DE MAR DEL PLATA.

pcgimene@mdp.edu.ar

Un modelo de regresión con errores de medición substanciales en las variables, principalmente en las covariables es conocido en la literatura como un modelo con errores en las variables (MEV). El caso funcional es obtenido cuando las verdaderas covariables son consideradas constantes fijas desconocidas o parámetros incidentales, cuyo número aumenta con el tamaño muestral. Los enfoques usuales de estimación de los parámetros de regresión o parámetros estructurales del modelo, tales como métodos de momentos, máxima verosimilitud, escore condicional o escore corregido entre otros, producen estimadores altamente sensibles ante contaminación de los datos y/o violación de los supuestos del modelo. Luego, métodos robustos en MEV funcionales son necesarios en varias aplicaciones. La literatura en este campo es escasa debido a la dificultad que acarrea la estimación en presencia de parámetros incidentales. En este trabajo proponemos un procedimiento de estimación para un MEV funcional lineal normal a partir del uso de una medida de divergencia cuasi logarítmica o q -divergencia, dependiendo de un parámetro real q . La motivación del enfoque se basa en la cuantificación de la discrepancia entre la densidad obtenida de los datos y la densidad del modelo. El estimador del parámetro de interés es obtenido minimizando esta divergencia como función de este parámetro desconocido. El método resulta equivalente a la minimización de una versión empírica de la entropía de Tsallis-Havdra-Charvat o a la maximización de la L_q -verosimilitud, siendo totalmente paramétrico y evitando los problemas que acarrea la estimación no paramétrica de densidades. La consistencia de Fisher es preservada a partir de una transformación para reescalar el parámetro. El procedimiento propone reemplazar los parámetros incidentales por estimadores que dependen de los parámetros estructurales. La constante q , con $0 < q < 1$, controla el balance entre eficiencia y robustez. El estimador de máxima verosimilitud es obtenido como caso

particular para $q=1$. Como estrategia de selección del parámetro q se considera la minimización del error cuadrático medio calculado por bootstrap paramétrico. Resultados asintóticos y propiedades de robustez de los estimadores son presentados. Su comportamiento es analizado además mediante estudios de simulación bajo diferentes escenarios de contaminación, elecciones de q y varianza del error de medición.

Palabras claves: modelos con errores; parámetros incidenta; estimación robusta; q -divergencia; l_q -verosimilitud

ESTRATEGIAS METODOLÓGICAS PARA DATOS DE PANEL EL CASO DE LOS BANCOS TÍPICOS EN ARGENTINA

MARGARITA DÍAZ y JOSÉ VARGAS

UNIVERSIDAD NACIONAL DE CÓRDOBA, FACULTAD DE CIENCIAS ECONÓMICAS.

mdiazlujan@gmail.com

El objetivo de este trabajo es analizar los efectos de indicadores específicos de los bancos en su rentabilidad, analizando para ello la información financiera de los bancos Típicos en Argentina para el período 2005-2015. En una primera etapa se clasificó a los bancos en diversos clusters mediante la aplicación del algoritmo K Medias robusto. A través del mismo se identificaron 4 clusters, caracterizados como Típicos, Otros Bancos, Comerciales y Personales. En este estudio se estimó el modelo para el grupo de 25 bancos considerados Típicos, de los cuales se excluyó uno en la estimación del modelo, dado su comportamiento atípico, detectado a través de un análisis exploratorio de los datos. La rentabilidad está medida a través del retorno sobre activo (ROA) y las variables predictoras son: Capital (Patrimonio Neto/Activo), Riesgo Crediticio (Previsiones/Préstamos), Productividad (Ingresos Brutos/Gastos Administrativos), Gestión de Gastos (Gastos Administrativos/Activo) y Tamaño (logaritmo del Activo). En una primera etapa se planteó un Modelo de Regresión Lineal que fue estimado utilizando Ordinary Least Squares (OLS), lo que permitió detectar observaciones influyentes. En una segunda etapa se trabajó con los métodos que consideran la estructura jerárquica de la información, esto es: Modelo de Efectos Aleatorios y Modelo de Efectos Fijos, con errores estándares robustos y con la corrección de Driscall and Kraay. A fin de seleccionar el método de estimación apropiado, se realizaron las pruebas de Hausman, de heterocedasticidad y de correlación transversal entre bancos. Los resultados indican que el modelo de efectos fijos con errores estándares con la corrección de Driscall and Kraay es el más apropiado para la estimación del modelo. Todos los determinantes, con la excepción de Gestión de Gastos, afectan significativamente la rentabilidad.

Palabras claves: rentabilidad de bancos; datos de panel; driscall and kraay; errores estándar robustos

ESTRUCTURA DE COVARIANZA ESTIMADA PARA EL MODELO ESTRUCTURAL ASIMÉTRICO DE CAUDALES DE DRENAJE

*NÉLIDA SUSANA OZÁN¹, CLAUDIA DE LOS RÍOS², GRACIELA GARCÍA¹,
EMMA MORALES¹, MARIANA GÓMEZ¹ y MARIO FERNENDEZ¹*

¹FACULTAD DE INGENIERÍA. UNIVERSIDAD NACIONAL DE SAN JUAN. ²FACULTAD DE CIENCIAS EXACTAS FÍSICAS Y NATURALES.
UNIVERSIDAD NACIONAL DE SAN JUAN.

sozan@unsj.edu.ar

La estimación del caudal de drenaje de una cuenca es de gran importancia, en la instancia del diseño, cuando se realiza la construcción de caminos de montaña. Una forma de estimar los caudales en la literatura tradicional de la ingeniería vial es utilizar el Método Racional, que no considera el aspecto aleatorio del fenómeno. Estudios previos muestran que las variables de mayor influencia en la determinación de caudales, por ejemplo la intensidad de las precipitaciones fluviales, son claramente sesgadas. Por este motivo, el ajuste del modelo estructural es una alternativa viable para este caso, asumiendo una distribución asimétrica (skew normal) para los errores. Se ha considerado la representación estocástica propuesta por Yuan y Bentler con la denominación de distribuciones asimétricas. La ventaja de esta forma de representación para las distribuciones permite modelar específicamente la estructura de covarianzas tanto del modelo estructural aditivo como multiplicativo, según corresponda. Por lo expuesto, en este trabajo, se estima la matriz de varianzas-covarianzas de resultados asintóticos para la distribución de los estimadores de los parámetros del modelo estructural asimétrico aditivo (Arellano-Valle, Ozán y Bolfarine, 2005). Esto permite probar la significancia de los parámetros del modelo y calcular los errores de estimación cometidos. El área de estudio donde se recolectó la información es el sector frontal de la Cordillera de Los Andes, San Juan (Argentina), contando con la colaboración de la Escuela de Caminos de Montaña (UNSJ). Los resultados muestran que, a pesar de haber asumido condiciones de regularidad para evitar problema de identificabilidad del modelo, se disminuyen los errores de estimación y se pueden plantear estudios de eficiencia asintótica.

Palabras claves: estructura de covari; modelo estructural; distribución asimétr

ÍNDICES DE DESARROLLO PSICOMOTOR POR PROVINCIA EN NIÑOS MENORES DE SEIS AÑOS. DATOS DE LA ENCUESTA NACIONAL DE SALUD 2005

DIANA M KELMANSKY¹, HORACIO LEJARRAGA², ALICIA MASAUTIS³ y FERNANDO NUNES⁴

¹INSTITUTO DE CALCULO, UBA. ²UBA. ³MINISTERIO DE EDUCACIÓN. ⁴AUTORIDAD CUENCA MATANZA RIACHUELO (ACUMAR).

dkelmansky@gmail.com

Se calculó un índice de desarrollo psicomotor (IDP) para cada provincia en base a datos de la encuesta Nacional de Nutrición y Salud (ENNys) realizada en 2005. El tamaño muestral con información sobre pautas de desarrollo fue de 13,323 niños (6536 varones). Para cada provincia se estimó la mediana de la edad de cumplimiento de cada pauta ajustando una regresión logística con la variable de respuesta cumple /no cumple y edad cronológica como variable explicativa. El IDP fue calculado sobre la edad de cumplimiento de diez pautas de desarrollo utilizando la recta estimada $y = a + bx$, donde y es: la diferencia entre la mediana de la edad de cumplimiento para la Referencia Nacional (x) y la mediana de la edad cumplimiento de una pauta. Para facilitar la interpretación de la pendiente el IDP fue calculado $IDP = 100*(1 + b)$. El valor teórico esperable de la pendiente es 0 si los niños no difieren en su desarrollo con los de la Referencia Nacional. En ese caso el IDP = 100. Se calculó el IDP para cada provincia; el rango fue entre 72.1 en el Chaco, hasta 106.4 en Tierra del Fuego. En la mayoría de las provincias los coeficientes de regresión estimados fueron negativos, indicando un aumento progresivo del retraso en la edad de cumplimiento de pautas con la edad de los niños. El coeficiente de correlación entre el IDP por provincia y la mortalidad infantil en 2005 fue extremadamente alto: -0.85, lo que sugiere que ambos indicadores comparten similares determinantes bio- sociales. El signo es negativo porque cuanto mayor es la mortalidad, menor es el IDP. El país dispone ahora de un indicador positivo de salud: el Índice de Desarrollo Psicomotor, simple de recoger, confiable y de bajo costo para ser incorporado a las estadísticas nacionales de salud.

Palabras claves: indicador; desarrollo psicomoto; regresión

MODELACIÓN DEL COMPORTAMIENTO ASIMETRICO DE LOS COSTOS EN EMPRESAS ARGENTINAS INCORPORANDO CASOS ATIPICOS CON GAMLSS

MARIA INES STIMOLO y JOSE VARGAS

FACULTAD DE CIENCIAS ECONOMICAS.

mstimolo@eco.unc.edu.ar

La función de costos de las empresas supone una proporcionalidad simétrica entre los costos y el nivel de ventas, es decir que los costos aumentan y disminuyen en la misma proporción ante el mismo cambio del nivel de actividad. Sin embargo, esta proporcionalidad no se cumple en la práctica y una de estas asimetrías, denominadas costos peggadizos, se manifiesta cuando los costos disminuyen en menor proporción que lo que se hubieran incrementado ante un aumento en el nivel de actividad. Anderson et.al. (2003) propone un modelo log lineal por partes para medir este comportamiento en las empresas y es muy común encontrar valores atípicos. Los trabajos empíricos sobre la problemática han eliminado esos valores extremos considerando distintos criterios. Aunque esos criterios pueden fallar cuando existe enmascaramiento de los mismos, y en consecuencia la estimación resultante no es apropiada. Ello se advierte cuando se realiza el diagnóstico de los residuos. Así, a partir de una muestra de empresas argentinas que realizan oferta pública de sus acciones durante el período 2004-2012, se estima el comportamiento de los costos utilizando modelos de regresión robusta y GAMLSS (Generalized Additive Models for Location, Scale and Shape) y se comparan parámetros estimados y empresas identificadas como atípicas proponiendo alternativas para encontrar un modelo que las logrando una mejor estimación. El modelo de regresión robusta identifica valores atípicos, que no difieren significativamente de los encontrados por los modelos tradicionales de detección a partir de las medidas de análisis de residuos del modelo de regresión, la estimaciones encontradas con GAMLSS resultaron superadoras ya que los valores considerados como atípicos en modelos basados en normalidad no resultaron tal cuando se consideraron otras distribuciones.

Palabras claves: modelos gamlss; regresión robusta; costos asimétricos

Muestreo

CÁLCULO DE PROBABILIDADES DE INCLUSIÓN PARA MUESTREO SISTEMÁTICO CON PROBABILIDAD PROPORCIONAL AL TAMAÑO CON REEMPLAZO DE UNIDADES

*GONZALO MARI¹, LUCÍA HERNÁNDEZ¹, GERARDO MITAS² y GABRIELA
BARBARÁ³*

¹INSTITUTO DE INVESTIGACIONES TEÓRICAS Y APLICADAS DE LA ESCUELA DE ESTADÍSTICA, UNIVERSIDAD NACIONAL DE ROSARIO.

²DIRECCIÓN NACIONAL DE METODOLOGÍA ESTADÍSTICA, INDEC. ³COORDINACIÓN DE MUESTREO, INDEC.

mari.gonzalo@gmail.com

La Muestra Maestra Urbana de Viviendas de la República Argentina (MMUVRA) es un instrumento a partir del cual es posible seleccionar muestras para los diversos operativos que se llevan a cabo en el Instituto Nacional de Estadística y Censos (INDEC). El diseño muestral consiste en uno con 2 etapas de selección. En la primera de ellas, las unidades primarias de muestreo están constituidas por los aglomerados urbanos que son seleccionados con un muestreo sistemático con probabilidad proporcional a la población total. En cada uno de los aglomerados seleccionados, se definen las unidades secundarias de muestreo como áreas, que en la mayoría de los casos equivalen a un radio censal y se seleccionan con un muestreo sistemático con probabilidad proporcional a la cantidad de viviendas del área. Cada una de las áreas seleccionadas son listadas y las viviendas que forman parte de las mismas constituyen el MMUVRA a partir de la cual se pueden seleccionar muestras para diversos operativos. Para cada uno de ellos se seleccionan viviendas, evitando la repetición de las mismas para no ocasionar un cansancio por parte de los respondentes que produciría un aumento en la no respuesta. Ante el agotamiento de una de las áreas por haber utilizado todas las viviendas, una de las soluciones que es la que utiliza el INDEC es reemplazar el área agotada por una nueva. Esto posee el inconveniente que la selección del reemplazo rompe con el diseño muestral planteado en la primera selección. Bajo esta situación, el cálculo de las probabilidades de inclusión de primer orden de las áreas seleccionadas resulta ser difícil. Una las soluciones posibles es realizar dicho cálculo a través de simulaciones. Thompson y Wu (2008) proponen un método basado en simulaciones que permite calcular las probabilidades de inclusión ante la ocurrencia de reemplazos para muestreo sistemático aleatorio con probabilidad proporcional al tamaño (PPT). El mismo se basa en la repetición del método de selección y del esquema de reemplazo un número grande de veces (por ejemplo, 1×10^6) y en base a los mismos, calcular las probabilidades de inclusión. La diferencia en el caso de la MMUVRA, es que el método de selección no es sistemático aleatorio PPT, sino sistemático PPT, el cual posee soporte mínimo (Pea, Qualité, Tillé, 2007). Se plantea una solución similar que se ajuste a la situación actual, y se consideran estudios futuros con soluciones alternativas.

Palabras claves: muestreo ppt sistemá; mmuvra; probabilidades de in; reemplazo de unidad

ENCUESTA DE ACTIVIDADES DE NIÑOS, NIÑAS Y ADOLESCENTES – RURAL (EANNA- RURAL) – DISEÑO MUESTRAL

***HUGO DELFINO¹, ANA MARÍA CATALANO², GERMAN ROSATI² y JOSÉ
ANCHORENA³***

¹UNLU. ²SUBSECRETARIA DE POLÍTICAS, ESTADÍSTICAS Y ESTUDIOS LABORALES. MTEYSS. ³SUBSECRETARIA DE POLÍTICAS, ESTADÍSTICAS Y ESTUDIOS LABORALES.MTEYSS.

hugo.delfino@dataexplorer.com.ar

En el año 2017, el MTEySS diseñó la muestra para la Encuesta de Actividades de Niños, Niñas y Adolescentes (EANNA) Rural. El objetivo de la EANNA Rural es analizar las actividades económicas y no económicas que realizan los niños/as y adolescentes de entre 5 y 17 años que habitan en viviendas particulares de zonas rurales, enfocando en las características, los contextos y los principales condicionantes del trabajo infantil. Desde un punto de vista de estadístico es un estudio basado en una muestra probabilística compleja. Se utilizó información de radios censales provistos por INDEC. Como la incidencia de la población rural dispersa varía según la región, se definieron dos estratos: a) radios con población agrupada y b) radios con población dispersa. Teóricamente corresponde a un: Diseño Estratificado de conglomerados a dos o tres etapas, con probabilidades variables de selección de las unidades de primera etapa (UPE) y probabilidad igual para las de segunda etapa (USE) en el caso de radios agrupados y variable en el caso de los radios dispersos. Se seleccionaron como UPE a radios dentro de cada estrato-dominio regional. Las USE

fueron en el caso de los radios agrupados una muestra de viviendas presuntamente habitadas, en los radios dispersos, dada su alta dispersión en la mayoría de los casos y debido a que no se dispone de un marco de muestreo ni es posible construir dicho marco, se procedió a marcar posibles viviendas (techos) utilizando sistemas de información geográfica e imágenes satelitales de Google Earth y Bing. En una etapa posterior, estos “techos” fueron clusterizados dentro de cada radio, para ello se buscó el número óptimo de clusters utilizando el Gap Statistic, luego utilizando el número óptimo del paso 1 se construyen los segmentos utilizando las coordenadas (X,Y) de los techos por medio del algoritmo K-means. y se seleccionaron dos de dichos clusters para, en una tercera etapa (UTE) seleccionar viviendas con probabilidades iguales. En los radios dispersos, se realiza un recorrido del área seleccionada para identificar si los “techos”, corresponden realmente a viviendas habitadas y detectar posibles omisiones, luego de actualizado el marco se procede a seleccionar viviendas de manera sistemática. La presente metodología muestra una alternativa novedosa para trabajar con áreas donde no es posible construir un marco de muestreo de manera tradicional. Los resultados suministraran información fehaciente a nivel nacional y regional que podrá ser utilizada para la formulación políticas de protección social basadas en evidencia.

Palabras claves: radios con población; radios con población; sistema de informaci; imágenes satelitales; trabajo infantil; población rural

MUESTRAS ALEATORIAS CON UNIDADES MUESTRALES IMPUESTAS: ¿ QUÉ HACER ?

ERNESTO A ROSA

UNTREF - DEPARTAMENTO DE METODOLOGÍA, ESTADÍSTICA Y MATEMÁTICA.

ear@fibertel.com.ar

Con el objeto de analizar el funcionamiento del sistema público sanitario argentino, un organismo encargado de la promoción y financiamiento del mismo, encargó la realización en todo el país de un estudio muestral complejo (es decir, con varias etapas de muestreo), destinado a cubrir una variedad de unidades de muestreo diferentes (centros de salud, sus profesionales y empleados, los usuarios, etc.), es decir con objetivos múltiples. Finalizado el Diseño Muestral y seleccionadas las unidades básicas del estudio (los centros de salud), pero antes de iniciarse el “trabajo de campo”, algunos representantes provinciales del Ministerio de Salud, consideraron que, debido a su conocimiento del sistema sanitario provincial, estaban en condiciones de opinar en relación a los centros de salud seleccionados aleatoriamente para la Muestra, y proponer modificaciones en los que debían ser finalmente encuestados. Naturalmente el Muestrista se negó rotundamente a realizar modificaciones en los centros seleccionados, argumentando en relación a la aleatoriedad utilizada para hacerlo, las bondades y el respaldo teórico de los métodos basados en el azar, y todos los fundamentos provistos por la Inferencia Estadística, pero los funcionarios nacionales eran permeables a las presiones provinciales, y no estaban dispuestos a perder el apoyo de personas “expertas e influyentes” en las provincias, por mantener los “caprichos” teóricos del Muestrista contratado. Esta imposición, que influía de forma diferente en las diversas provincias cubiertas, impulsó al Muestrista a convocar a otros profesionales del Muestreo, ante los cuales expuso el problema, debatieron sobre el mismo, se rebelaron rotundamente contra las influencias externas, pero encontraron una forma simple de mantener la “pureza” de un Diseño Muestral aleatorio, con los compromisos de los funcionarios del organismo central, y los antojos y preferencias de los personajes provinciales: considerar a los centros de salud “impuestos”, como casos de inclusión forzosa (que ya existían por otras razones), y excluirlos de la población de referencia representada con la Muestra aleatoria. En este trabajo se expone sucintamente la forma en que se enfrentaron las imposiciones de los funcionarios, para lograr tener una muestra probabilística, con el único problema conceptual de tener que modificar los factores de expansión de cada una de las unidades seleccionadas al azar, para que representen a

la población finalmente considerada.

Palabras claves: muestreo; diseño muestral; muestras de objetivo; inferencia estadística

MUESTREO DE POBLACIONES BERNOLLI: RELACIÓN ENTRE PARÁMETROS POBLACIONALES Y MUESTRALES EN PRESENCIA DE FUENTES DE VARIABILIDAD ADICIONALES AL MUESTREO

VIVIANA B LENCINA

INSTITUTO DE INVESTIGACIONES ESTADÍSTICAS, FACULTAD DE CIENCIAS ECONÓMICAS, UNT.

lencina.viviana@yahoo.com.ar

En un muestreo probabilístico los parámetros de las variables observadas (muestra) se relacionan con los parámetros de la población muestreada, y es así como se usa la información de la muestra para inferir sobre la población. En este trabajo se aborda el problema de Inferencia en Poblaciones Bernoulli, se extiende la aplicación de la distribución binomial a situaciones donde no son idénticamente distribuidas las variables Bernoulli en la población objetivo y se muestra cómo se relacionan los parámetros poblacionales con los muestrales ante la presencia de errores de respuestas endógenos y exógenos. Por población Bernoulli en presencia de errores se entiende una población que posee una característica dicotómica que se presenta de manera aleatoria, errores de respuesta endógeno, o una característica dicotómica fija, pero que al introducirse un error exógeno en el proceso de medición se modifica su identificación. Específicamente se presenta cómo, en el proceso de muestreo, las diferentes fuentes de variabilidad involucradas, variabilidad intra individuo (errores de respuesta endógenos), inter individuo (errores de respuesta exógenos), o del proceso probabilístico para realizar el muestreo (errores muestrales) combinan los parámetros de las variables aleatorias que componen la población para obtener los parámetros de las variables realizadas en la muestra. Se describen situaciones prácticas donde queda justificada la utilización de la distribución binomial aún cuando el clásico modelo de urna no pueda aplicarse. Es importante resaltar que el análisis de la relación entre los parámetros muestrales y poblacionales es fundamental para poder especificar operacionalmente los objetivos de cualquier problema aplicado y seleccionar la estrategia de inferencia estadística a ser utilizada.

Palabras claves: población bernoulli; errores de respuesta; muestreo

MUESTREO ESTRATIFICADO Y ENFOQUE DE MODELOS PARA ESTIMAR TOTALES CON DATOS ESPACIALES EN ESTUDIOS SOCIOECONÓMICOS

VIRGINIA LAURA BORRA, CÉSAR ANTONIO MIGNONI y JOSÉ ALBERTO PAGURA

FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA-UNIVERSIDAD NACIONAL DE ROSARIO.

vborra@fcecon.unr.edu.ar

En los estudios por muestreo, es frecuente encontrarse con la necesidad de realizar inferencias en poblaciones cuyas unidades se encuentran ubicadas en el espacio. En esta situación, por lo general, se observará la presencia de autocorrelación espacial y el plan de muestreo puede volverse más eficiente si se tiene en cuenta esta característica. En estos casos, se ha propuesto la selección de la

muestra mediante muestreo sistemático o muestreo estratificado y podrán lograrse importantes mejoras en la precisión de las estimaciones si se emplea la información proporcionada por un modelo que exprese dicha autocorrelación espacial (modelo de semivariograma) en dicha selección. Diferentes autores han presentado expresiones matemáticas para evaluar la eficiencia de los métodos de selección mencionados con respecto al muestreo aleatorio simple como funciones de la correlación espacial. Por otra parte, la información contenida en el semivariograma, también puede utilizarse en el proceso de estimación recurriendo al enfoque inferencial basado en modelos, el cual basa la estimación de los valores poblacionales en la predicción de los valores de la variable, en las unidades no incluidas en la muestra, por medio de la utilización de un modelo teórico. Este modelo puede contener variables auxiliares relacionadas con la variable en estudio, así como la información proporcionada por el semivariograma. Como medida de precisión, se estima el error cuadrático medio del predictor en la superpoblación. Borra (2015), Pagura y otros (2015) presentan la aplicación de los métodos de selección mencionados en estudios socioeconómicos, mostrando mejor eficiencia en los estimadores de simple expansión con respecto al muestreo aleatorio simple, considerando el error cuadrático medio en la población finita (enfoque de diseño). Además muestran que el empleo del predictor utilizando modelos de semivariograma, produce mejoras en el error cuadrático medio para la población finita (enfoque asistido por modelos). Por otra parte, la estratificación espacial permite definir grupos homogéneos, sin embargo las variables que se estudian podrían aun presentar correlación espacial dentro de cada estrato. Por este motivo, puede plantearse el uso de semivariogramas para obtener, con el enfoque de modelos, estimaciones separadas o combinadas de totales y promedios. En este trabajo se presentan los resultados de esta propuesta para la estimación del total de hogares con necesidades básicas insatisfechas en la ciudad de Rosario en el 2010, tomando como unidad muestral el radio censal, y se aprovecha la disponibilidad de la información censal, para la realización de comparaciones.

Palabras claves: datos espaciales; muestreo estratifica; enfoque de modelos; modelos de semivario

POBLACIONES OCULTAS: KIDS ONLINE, MUESTRA DE HOGARES CON MARCO IMPERFECTO

JUAN JOSÉ GOYENECHE

FACULTAD DE CIENCIAS ECONÓMICAS.

jjgoye@iest.a.edu.uy

En muestreo los marcos imperfectos son la regla más que la excepción. En general hay un cierto grado de desajuste entre el marco y la población de estudio, lo que lleva a temas de sub- o sobre-cobertura, o problemas en las probabilidades de selección. La ausencia de marcos de lista nos lleva a usar alternativas para el estudio de poblaciones ocultas. Kids Online (Niños Conectados, en español) es un estudio realizado en varios países. El objetivo es estudiar como niños entre 9 y 17 años se relacionan con Internet, como y para que la usan, desde que dispositivos, los riesgos asociados con el uso y que oportunidades les da el acceso a la red. “El estudio Kids Online fue realizado por primera vez en países de la Unión Europea en el año 2006 ante la iniciativa de un equipo de académicos del London School of Economics (LSE). A la fecha continúa realizándose de manera regular en varios continentes. En América del Sur lo han realizado Brasil, Chile y Argentina, y este año por primera vez Uruguay” (tomado de la carta de presentación a las familias seleccionadas) En el caso que se presenta se desean estudiar hogares con niños en las edades de 9 a 17 años (Hogares KO). Estos hogares KO son parte de los incluidos en el Censo (aproximadamente un 25% de los hogares tienen niños en esas edades), pero no se dispone de sus ubicaciones actuales. Se desarrolla un mecanismo para ubicar estos Hogares KO y se estudia la eficiencia del procedimiento y los posibles sesgos que implique. El procedimiento diseñado implica que las probabilidades de inclusión de los

Hogares KO son desconocidas y deben ser aproximadas a partir de información censal. Palabras clave: poblaciones ocultas, probabilidades desiguales de inclusión

Palabras claves: poblaciones ocultas; probabilidad desigual; estudio kids online

Probabilidad y procesos estocásticos

GESTIÓN EFICIENTE DE RIESGOS FINANCIEROS UTILIZANDO UN MODELO ESTOCÁSTICO DE DOS FACTORES

MARÍA TERESA CASPERRI, MARTÍN EZEQUIEL MASCI y JOAQUÍN BOSANO

UNIVERSIDAD DE BUENOS AIRES, FACULTAD DE CIENCIAS ECONÓMICAS. IADCOM, CENTRO DE INVESTIGACIÓN EN MÉTODOS CUANTITATIVOS APLICADOS A LA ECONOMÍA Y LA GESTIÓN..

martinmasci@economics.uba.ar

Los instrumentos derivados financieros están sujetos a condiciones de incertidumbre que justifican la importancia de tener modelos de valuación que sean a la vez precisos y parsimoniosos. En la actualidad existen un gran número de ellos, por lo tanto, resulta importante disponer de un criterio de validación exploratorio para asistir al analista en el proceso de selección del mejor modelo. En este sentido, la eficiencia del modelo debe basarse en la decisión financiera asociada y las características intrínsecas de la herramienta. Respecto de la primera característica, el uso de modelos estocásticos en ciencias económicas representa una herramienta para el análisis crítico de las decisiones. Por otro lado, las características intrínsecas de dichos modelos estadísticos facilitan la implementación en términos de velocidad de procesamiento y parsimonia de las variables involucradas. Este trabajo implementa un modelo canónico de valuación de futuros sobre petróleo crudo -Gibson-Schwartz (1990)- y busca validar los estimadores de sus parámetros para explorar su precisión y eficiencia. Dada la naturaleza de los datos, el método de estimación elegido es un modelo de seemingly unrelated regression (SUR), una generalización del modelo de regresión lineal multiecuacional que permite que los residuos estén correlacionados. Para hacer esto, se desarrolla un proceso de validación basado en casos, implementado en un código de STATA. Dichos casos contemplan la valuación del precio de los derivados, en distintas ventanas temporales. De esta manera, el grado de eficiencia que se le exige al modelo se basa en la parsimonia del mismo, a la vez que identifica las dificultades de cada variable en términos de su bondad del ajuste. Esta investigación concluye que el modelo propuesto es satisfactorio, dada su naturaleza parsimoniosa, pero existe lugar para mejoras si se consideran algunos modelos alternativos.

Palabras claves: procesos estocástico; modelos multi-factor; derivados financiero; futuros

Series de tiempo

EX-ANTE VOLATILITY MEASURES BASED ON AGENTS' EXPECTATIONS AN APPLICATION TO ARGENTINA'S TERMS OF TRADE

SERGIO MARTÍN BUZZI¹ y JOSÉ LUÍS ARRUFAT²

¹FCE - UNC. ²FCE -UNC.

sergiomartinbuzzi@gmail.com

In this paper we discuss the construction of volatility measures aimed to reflect the ex-ante uncertainty faced by economic agents. The most widely used volatility measure is the standard deviation of the original time series. This measure no doubt has some merit given that the future values of a series with high variability is probably more difficult to predict than future values from a series which is less variable. But, the raw standard deviation implicitly assumes that economic agents cannot anticipate any component of the series. However, if agents perceive a portion of the data generation process, they are surprised only by the unpredictable component. Following, volatility measures based in the unpredictable components can be built in order to obtain a better proxy for agents' true uncertainty. In order to do that, volatility measures are computed using the residuals from a set of predictive models. Each one of those models can be viewed as an alternative assumptions made about the expectation formulation process faced by economic agents. Then, assuming different models of expectation formulation, volatility indices are built removing the components of total fluctuations that are predictable, and measuring the variability of the unpredictable component. Almost all the previous literature following this line, computes volatility using the residuals from a predictive equation estimated for the full sample period under study; but as we noted in Arrufat et al. (2014), economic agents cannot base their forecasts using data from the unseen future, that is, the predictive model only can be based on the data available at the moment when the forecast is made. A practical procedure to avoid the latter problem is to implement a rolling window estimation algorithm, on which agents formulate their predictions for next period based upon the k most recent periods. Finally, the alternative indices built are employed to compute the volatility of the Terms of Trade (TOT) of Argentina and the estimates are compared in order to conclude if they follow different paths.

Palabras claves: volatility; uncertainty; rolling window; time series

LA HETEROCEDASTICIDAD CONDICIONAL EN LA INFLACIÓN DE LA ARGENTINA (1943 - 2013)

MARÍA DE LAS MERCEDES ABRIL y JUAN CARLOS ABRIL

UNIVERSIDAD NACIONAL DE TUCUMÁN. CONICET.

mabrilblanco@hotmail.com

Numerosas series de tiempo económicas no tienen una media constante y en situaciones prácticas, frecuentemente vemos que la varianza del error observacional está sujeta a una sustancial variabilidad a través del tiempo. Ese fenómeno es conocido como volatilidad. Para tomar en cuenta la presencia de la volatilidad en una serie económica es necesario recurrir a modelos conocidos como modelos heterocedásticos condicionales. En estos modelos, la varianza de una serie en un dado

instante de tiempo, depende de la información pasada y de otros datos disponibles hasta aquel instante de tiempo, de modo que se debe definir una varianza condicional, que no es constante y no coincide con la varianza global de la serie observada. Existe una variedad muy grande de modelos no lineales en la literatura, útiles para el análisis de series de tiempo económicas con volatilidad, pero nos concentraremos para analizar la serie de interés, en los modelos de tipo ARCH introducidos por Engle (1982) y sus extensiones. Estos modelos son no lineales en lo que se refiere a la varianza. Nuestro objetivo será el estudio de los datos mensuales de inflación de la Argentina para el período enero de 1943 a diciembre de 2013. Los datos son los publicados oficialmente por Instituto Nacional de Estadística y Censos (INDEC). Si bien es un período muy largo en el cual sucedieron cambios de base, cambios de canasta e intervenciones en el INDEC, se puede apreciar que ciertos patrones generales de comportamiento han persistido en el tiempo, lo que nos permite admitir que el estudio está adecuadamente basado en la información disponible.

Palabras claves: inflación; heterocedasticidad; volatilidad; series de tiempo

ON THE ROBUSTNESS OF THE PRINCIPAL VOLATILITY COMPONENTES

CARLOS TRUCÍOS¹, LUIZ K HOTTÀ² y PEDRO VALLS¹

¹SÃO PAULO SCHOOL OF ECONOMICS, FGV - BRAZIL. ²UNIVERSITY OF CAMPINAS - BRAZIL.

carlos.trucios@fgv.br

In this work, we analyze the effects of additive outliers on the recently proposed procedure of Hu and Tsay (2014) (Principal volatility component analysis. Journal of Business & Economic Statistics, v32.2) and Li et al. (2016) (Modeling multivariate volatilities via latent common factors. Journal of Business & Economic Statistics, v34.4) called principal volatility components. This procedure overcome several difficulties in modeling and forecasting the conditional covariance matrix in large dimensions arising from the curse of dimensionality. We show that outliers have a devastating effect on the construction of the principal volatility components, in the estimation of the number of components and on the forecast of the conditional covariance matrix in both small and large dimensions and in consequence in economic and financial applications. In fact, we show that just a few outliers are necessary to affect drastically the volatility components and the estimation of the conditional covariance matrix. We propose a robust procedure based on a robust estimator of the unconditional covariance matrix, on a weighted estimator that capture the volatility dynamic and in a robust filter. We analyze the finite sample properties by means of Monte Carlo experiments considering several patterns of contamination and show that the robust procedure outperforms the classical method in contaminated series and has a similar performance in uncontaminated ones. The new procedure is also applied to 73 stock returns of the Nasdaq 100 index and a minimum variance portfolio is constructed. Results are compared with the non-robust version. The new procedure showed a better performance in both simulated and real data.

Palabras claves: covariance matrix; forecast; high-dimensional data; principal components; outliers; volatility

UNIT ROOT TESTING UNDER STRUCTURAL BREAKS

SERGIO MARTÍN BUZZI¹ y SILVIA MARÍA OJEDA²

¹FCE - UNC. ²FAMAF.

sergiomartinbuzzi@gmail.com

In this paper are analyzed eleven stock market indices in order to conclude about their integration orders. With this objective in mind, three tests are performed. The first test, is the standard Augmented Dickey-Fuller (ADF) unit root test. It is known that this test can face problems of lack of power, failing to reject the null hypothesis being the series in fact integrated of order zero, I(0), erroneously concluding the existence of unit roots. The second test, is the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) stationarity test, which complements the ADF test, provided that its null hypothesis is that the series under analysis is stationary. The third test posed is the Kapetanios unit root test, which is an extension of Zivot and Andrews' unit root test for the case of up to m structural breaks. This third test is intended to solve another problem faced by standard ADF test which could conclude the existence of a unit root, when in fact the series is integrated of order zero with a broken deterministic tend. The estimations are performed using daily data for a long time period, for the nine greater world stock markets indices plus Bovespa and Merval indices. The testing procedures are run in the open source statistical programming language R. Moreover, an R procedure is written in order to perform the Kapetanios test, modifying the existing ur.za function from urca package. Finally the results from those tests are compared and interpreted, reaching the conclusion that the series are integrated of order one, I(1).

Palabras claves: unit root; time series; structural breaks; stock markets

Teoría sobre distribuciones de probabilidad

BONDAD DE AJUSTE DE LA DISTRIBUCIÓN DE UN ÍNDICE DE EXPOSICIÓN A PLAGUICIDAS APLICACIÓN EN TRABAJADORES RURALES DE LA PROVINCIA DE CÓRDOBA

OLGA BEATRIZ PADRÓ¹ y MARÍA DEL PILAR DÍAZ²

¹FACULTAD DE CIENCIAS ECONÓMICAS- UNIVERSIDAD NACIONAL DE CÓRDOBA. ²INICSA-FACULTAD DE CIENCIAS MÉDICAS -UNIVERSIDAD NACIONAL DE CÓRDOBA.

olgapadro@gmail.com

El estudio de los problemas ambientales, productos de fenómenos que implican acumulación de algún tipo de tóxicos, o la concentración de contaminantes en el aire y en el agua (Vilca et al., 2010), como así también la exposición acumulada a plaguicidas que afecta a los trabajadores rurales, constituye un desafío analítico y estadístico. Estos fenómenos involucran variables de vida caracterizadas por ser asimétricas, unimodales, sesgadas positivamente y de dos parámetros, con momentos de cualquier orden (Marshall & Olkin, 2007). Tradicionalmente, distribuciones como la exponencial, Weibull, lognormal, gamma, Birnbaum-Saunders y Gaussiana inversa, son las comúnmente utilizadas a pesar de presentar bondad de ajuste deficiente en los eventos extremos o en las colas de la distribución. Para mejorar el ajuste diversos modelos de vida, como los modelos GBS (Generalized Birnbaum-Saunders) (Leiva et al. 2008) y los modelos IGT (Inverse Gaussian Type) (Sanhueza et al., 2008), han sido propuestos. Los primeros se caracterizan por admitir diferentes grados de curtosis y asimetría, así como unimodalidad y bimodalidad; los segundos son considerados robustos, desde el enfoque de Lange et al. (1989) ya que es una clase de modelos que incluye diferentes grados de curtosis. El presente trabajo aborda dos índices para la valoración de la exposición a plaguicidas: el Nivel de Intensidad a la Exposición (IE) y la Exposición Acumulada (EAC), construidos por el Grupo de Epidemiología Ambiental del Cáncer en Córdoba (GEACC, Lantieri et al., 2011). El objetivo fue definir el modelo probabilístico de mejor ajuste para cada uno, utilizando y

comparando las distribuciones tradicionales con los modelos GBS e IGT, y seleccionando los más adecuados según sus propiedades teóricas e interpretaciones. Con el fin de validar la distribución seleccionada como de mejor ajuste, se estudian los comportamientos de las tasas de riesgo de dichos índices (IE y EAC), utilizando TTT Plots (Aarset, 1987). Una vez elegido el modelo de distribución se obtienen los percentiles que determinan los niveles de baja, media y alta exposición a plaguicidas.

Palabras claves: índices de exposición; bondad de ajuste; distribuciones de vi; tasas de riesgo; ttt plot

Otras categorías metodológicas

ANÁLISIS DE UN PROYECTO ESTADÍSTICO BASADO EN UN SISTEMA DE INDICADORES DE RAZONAMIENTO Y PENSAMIENTO ESTADÍSTICO

MARIEL LOVATTO y LILIANA TAUBER

UNIVERSIDAD NACIONAL DEL LITORAL.

estadisticabiologiafhuc@gmail.com

En el presente trabajo describimos el significado de los indicadores que consideramos dentro de un sistema. Estos indicadores nos permiten analizar las relaciones entre conceptos estocásticos, que pueden ponerse en práctica cuando se elabora una propuesta didáctica basada en el trabajo por proyectos. Tomando como fundamento el sistema de indicadores mencionado antes, describimos el análisis de contenido realizado sobre un proyecto para el aula de Estadística, elaborado por un docente que ejerce actualmente en el Nivel Secundario. Por último, realizamos una discusión de los resultados obtenidos a partir del análisis de contenido, problematizando sobre los conflictos que encontramos en nuestro caso de estudio en relación con el razonamiento y pensamiento estadísticos. Para la elaboración de los indicadores nos basamos en las dimensiones uno y dos del pensamiento estadístico de Wild & Pfannkuch (1999), definidos como ciclo investigativo y tipos de pensamientos respectivamente. Dentro de la dimensión dos encontramos los cinco componentes que, según las autoras, son fundamentales para el desarrollo del pensamiento estadístico: Necesidad de los datos, Transnumeración, Variabilidad, utilización de modelos, contexto y conceptos. Dentro de cada componente definimos indicadores los cuales utilizamos para analizar un proyecto, resultado del Trabajo Final obligatorio para aprobar el Módulo de Enseñanza de la Probabilidad y la Estadística (Tauber e INFID, 2016), el cual es una instancia obligatoria en el cursado de la Especialización Docente de Nivel Superior en Enseñanza de la Matemática en la Educación Secundaria, ofrecida de manera virtual por el Instituto Nacional de Formación Docente (INFID). Logramos observar en el análisis del proyecto desajustes entre lo planteado por los docentes en cada actividad, objetivo y problemas, con respecto a los indicadores planteados. Se espera desarrollar análisis de nuevos proyectos para comparar entre ellos. Estos conflictos nos hacen reflexionar sobre la necesidad de promover la formación de los docentes que están encargados de alfabetizar estadísticamente a ciudadanos que puedan actuar críticamente y debatir de manera fundamentada ante la información estadística.

Palabras claves: razonamiento estadístico; pensamiento estadístico; evaluación continua; formación de profesores

CURVAS ROC COMO MEDIDA DE DIAGNÓSTICO EN MODELOS

SHRIKANT BANGDIWALA¹, ANA MARÍA SFER² y M A D'URSO VILLAR²

¹MCMASTER UNIVERSITY. ²UNIVERSIDAD NACIONAL DE TUCUMÁN.

asfer@herrera.unt.edu.ar

El ajuste de modelos lineales generalizados mixtos (GLMM) ha dado respuestas a una amplia gama de estudios observacionales y experimentales cuando las observaciones están correlacionadas. En estos modelos, el análisis de los residuos no tiene una discusión acabada en la bibliografía y no todos los paquetes disponibles tienen en cuenta el efecto aleatorio en las predicciones. El modelo más usado dentro de los GLMM es el de respuesta binaria. Para el análisis de los residuos, en este trabajo, se utiliza la curva ROC y su respectiva área bajo la curva (AUC), herramientas clásicas del modelo logístico. También se usa el coeficiente de concordancia Kappa de Cohen y el Test de McNemar. El análisis de la curva ROC, permite seleccionar los modelos óptimos independiente-mente de la distribución de éxitos y fracasos. EL AUC se usa para medir el poder de discriminación del modelo. El coeficiente Kappa de Cohen, permite ver si las predicciones clasifican bien a las observaciones. El Test de McNEmar compara las proporciones de correctamente clasificados por el modelo. Los objetivos son estudiar, mediante simulaciones, el comportamiento de las predicciones de GLMM con respuesta binaria teniendo en cuenta o no los efectos aleatorios, y compararlas con las correspondientes a las del modelo logístico. Se genera una variable (Y1,Y2,Y3,X1,X2) Normal Multivariada (100,0,). Se considera Yi como las respuestas en 3 repeticiones y Xi las covariables. Se categorizan las variables respuestas considerando punto de corte en cero, y una de las covaria-bles considerando punto de corte su mediana. Se ajusta un modelo logístico y un modelo logístico mixto. Se generan las predicciones en la escala link y de probabilidad en cada modelo. En el mo-delo mixto se generan con y sin efecto aleatorio. Se construyen las correspondientes curvas ROC. Los resultados indican que las mayores áreas bajo la curvas ROC las tiene el modelo mixto que incluye en la predicción los efectos aleatorios. La concordancia entre la respuesta observada y la predicción es fuerte, aproximadamente igual a 0.70, en estos modelos y moderada, próxima a 0.56, en el modelo mixto cuando no se incluyen los efectos aleatorios. Los resultados más pobres se obtienen con el modelo logístico (conceptualmente erróneo). Las pruebas de McNemar muestran resultados similares, aunque son mejores para el modelo mixto cuando ellas consideran los efectos aleatorios. También, se concluye que los resultados son mejores cuando se considera las repeticio-nes por separado.

Palabras claves: modelos mixtos; curvas roc; concordancia

EL ROOL DEL REFERENTE ESTADÍSTICO EN EL AREA DE SALUD PUBLICA

JUAN PABLO INCOCCIATI¹ y DIANA ROSANA KUCUKBEYAZ²

¹MUNICIPIO DE LA MATANZA. ²UNTREF.

jpabloinco@gmail.com

EL ROOL DEL REFERENTE ESTADISTICO EN EL AREA DE SALUD PUBLICA Autores: Juan Pablo Incocciati , Diana Rosana Kucukbeyaz mail: jpabloinco@gmail.com Resumen: ¿QUE ES UN LIDER ¿QUIEN ES UN REFERENTE Hay lideres formales que solamente ocupan un cargo, pero el liderazgo autentico es básicamente la sumisión de significados y valores que obran como líneas de fuerza, estimulantes en el corto plazo e irradiadas hacia el largo plazo. El líder no

tiene que ser excepcional, sino simplemente debe generar una relación excepcional con la organización que permite COMPROMETER más que involucrar. Los que ejecutan las tareas son los que mejor conocen los detalles del trabajo, sus posibles errores y las mejores alternativas para solucionarlo. Todo miembro de una organización está en condiciones y generalmente desea contribuir con la realización de una buena función. Las contribuciones proveen un sentido de pertenencia y favorecen la integración entre distintos niveles. La utilización de procesos estructurados para la mejora y resolución de los problemas usando medios estadísticos de control, produce a largo plazo mejores resultados que el control no estructurado, la incorporación de personal con formación estadística en los efectores de salud, genera un aporte de calidad y confianza a los datos y su tratamiento. Gerenciamiento, gestión y coordinación. Actitud de respeto y valoración a todas las actividades del área que se hicieran antes de su incorporación. Adaptación local de las propuestas acorde a la realidad de la región, respetando pautas culturales, sociales, económicas y políticas. Evaluación en forma permanente para producir reformas rápidas y acordes a los requerimientos locales. Facilitador de propuestas, utilizando a la comunicación como una herramienta de conducción participativa y abierta, dispuesto al aprendizaje continuo y a la innovación como respuesta a la variabilidad del sistema. Palabras claves Rol , Líder, Referente Eje temático: Salud Humana Conclusiones: Solucionar en parte la crisis de valores, colaborar en el descubrimiento de las verdaderas necesidades de la población. en el área de referencia, generar datos confiables que ayuden a la toma de decisiones en un área tan sensible como la salud pública. Palear a partir de esto las Fallas en la comunicación entre distintos niveles y con la comunidad. Utilización de un enfoque sistemático en el análisis de las situaciones

Palabras claves: rol; líder; referente

ESTUDIO DEL IMPACTO REDISTRIBUTIVO DE LA ASIGNACIÓN UNIVERSAL POR HIJO EN ENTRE RÍOS. UNA APLICACIÓN DE LA METODOLOGÍA DE DESCOMPOSICIÓN DEL COEFICIENTE DE GINI

STEFANIA D'ITORIO y OLGA B. AVILA

FACULTAD DE CIENCIAS ECONOMICAS - UNIVERSIDAD NACIONAL DE ENTRE RIOS.

stefaniaditorio@gmail.com

La desigualdad en la distribución del ingreso se ha convertido en un grave problema de los argentinos desde el comienzo de la última dictadura militar, en 1976. Este problema responde a la interacción entre distintos factores -diferencias salariales, niveles de empleo, instituciones laborales, políticas sociales, condiciones macroeconómicas, entre otros-. En este marco, los estudios de las políticas públicas sobre las distintas fuentes de ingreso y su impacto sobre la desigualdad resultan relevantes. Este trabajo tuvo como objetivo estimar el impacto distributivo de la Asignación Universal por Hijo (AUH) en la desigualdad del ingreso en los aglomerados de Gran Paraná y Concordia, mediante la aplicación de una metodología de descomposición del coeficiente de Gini por fuentes de ingresos, entre el 3er trimestre de 2009 -trimestre anterior a su puesta en vigencia-, hasta el 2do trimestre de 2015. Mediante la metodología de descomposición propuesta por Lerman y Yitzhaki (1985) se buscó ahondar en la desigualdad total, mediante el cálculo de la participación relativa de la fuente transferencias en el ingreso total, la desigualdad observada en dicha fuente, y la correlación entre ésta y el ingreso del hogar. También, con el cálculo de la elasticidad-Gini del ingreso propuesto por Wodon y Yitzhaki (2002), se evaluó cómo un cambio marginal en la composición del ingreso por transferencias modificó la desigualdad total, de manera de cuantificar el aporte de esta fuente en la distribución del Ingreso Per Cápita Familiar en la dinámica que este último ha seguido desde la implementación de la AUH. Esta investigación contribuyó al estudio de la desigualdad a nivel subprovincial, limitándose a cuantificar el aporte de las transferencias en la variación de la

desigualdad total y su evolución en los aglomerados mencionados, sin intenciones de establecer de manera definitiva las causas de los cambios en la desigualdad del ingreso. Se observó un aumento del impacto redistributivo de esta fuente en la desigualdad total en ambos aglomerados debido a la AUH: la variación en el coeficiente de Gini debido a las transferencias aumentó en el trimestre de su implementación, en los posteriores a cada aumento del monto, y entre las puntas del período analizado, con mayor énfasis en Concordia. Pero debido a su baja participación en el ingreso total se concluye que, en la medida en que los objetivos de este tipo de políticas busquen mejorar la distribución del ingreso, serán precisos mayores esfuerzos fiscales para incrementar el monto de la transferencia.

Palabras claves: impacto redistributi; asignación universal; metodología de desco

EVALUACIÓN DE IDEAS PARA PROMOVER EL COMPROMISO Y EL APRENDIZAJE DE LA ESTADÍSTICA

GABRIELA DAMILANO¹ y DAIANA RIGO²

¹UNIVERSIDAD NACIONAL DE RÍO CUARTO. ²UNIVERSIDAD NACIONAL DE RÍO CUARTO - CONICET.

gdamilano@gmail.com

Durante el período 2015-2016 desde la cátedra de Estadística en Ciencias Sociales se desarrolló una propuesta de innovación pedagógica con la finalidad de favorecer y promover el compromiso, la motivación y la comprensión de las herramientas básicas por parte de los estudiantes. La misma consistió en una configuración didáctica organizada en distintos ejes, en uno de los cuales los estudiantes desarrollaron un proyecto de investigación (TCI) original e inesperado, desde la formulación del problema hasta la presentación y discusión de resultados. En ambos años, la innovación se evaluó con un diseño longitudinal de grupo. El seguimiento del desempeño de los estudiantes se realizó, a través del rendimiento académico obtenido durante el cursado de la asignatura; la valoración del proceso y defensa del TCI junto a entrevistas semiestructuradas para indagar la participación y el compromiso con la propuesta; y la aplicación de un instrumento antes y después de la innovación, para saber qué conocimientos tienen los estudiantes y cómo los aplican para atender a la lectura e interpretación de información estadística. En el primer año, se consideró como instrumento de evaluación la proyección del video Desmontando mitos sobre el mundo (entrevista a Hans Rosling) y una serie de preguntas que buscaban que los estudiantes relacionaran el contenido de la charla con conceptos básicos de estadística: problema de investigación, población, muestra, unidad de análisis, variables y la interpretación de información estadística a través nociones centrales como tendencia, variabilidad y predicción. En el segundo año, con el fin de que comprendieran la importancia del contexto para interpretar los datos, se utilizó como disparador la actividad Papeles en la espalda, donde los estudiantes debían registrar las respuestas que sus compañeros brindaban a la pregunta que tenían pegada en su espalda; con la información recolectada debían comentar los aspectos más destacados, realizar un gráfico lo más creativamente posible y en función de ello intentar adivinar cuál era la pregunta formulada en su espalda. Además y para valorar la compresión de la información brindada, debían conectar a través de una línea, una serie de enunciados curiosos que circulan en la web con el gráfico estadístico que consideraran mejor representaba la información proporcionada y le permitiera responder a las preguntas asociadas. Los resultados muestran que aprender bajo propuestas pedagógicas innovadoras, originales y novedades no sólo favorece el interés y la participación sino que además contribuye a crear un contexto promisorio para el aprendizaje de las estadísticas.

Palabras claves: alfabetización estad; innovación; motivación; evaluación

EXPERIENCIA PEDAGÓGICA EN LA ENSEÑANZA DE ESTADÍSTICA APLICADA EN LA CARRERA DE RELACIONES DEL TRABAJO: EL ÁMBITO DEL TRABAJO Y TECNOLOGÍAS DE LA INFORMACIÓN

GRACIELA DURET

TITULAR ESTADÍSTICA APLICA- CARRERA RELACIONES DEL TRABAJO- F.CIENCIAS SOCIALES.UBA.

gcduret@gmail.com

Durante el período 2013-2015 se llevó a cabo en la Cátedra de Estadística Aplicada (I y II) de la Carrera de Relaciones del Trabajo, en el marco del Programa de Reconocimiento Institucional de Investigaciones, siendo la especialidad temática Estadística y Gestión del Conocimiento, el proyecto “El ámbito del trabajo y su integración con las Tecnologías de la Información y Comunicación (TIC)”. El objetivo general del proyecto fue introducir a los alumnos en la utilización de métodos estadísticos para llevar a cabo investigaciones relacionados al mundo laboral, dando los primeros pasos en ambos campos, el de la estadística y el de la investigación. Se plantearon dos desafíos: una experiencia pedagógica y llevar a cabo una investigación cuantitativa. El objetivo pedagógico fue explorar la factibilidad de formar equipos de investigación que involucraran alumnos de grado y realizar trabajos que continuasen después del cursado de las materias. El objetivo de la investigación cuantitativa fue relevar y describir nuevas formas de trabajar mediadas por las TIC. Dada la relevancia que tiene para los profesionales de Relaciones del Trabajo el conocer la problemática de la incidencia de las TIC en el mundo laboral, a pedido de los alumnos, se decidió investigar acerca del teletrabajo, y sobre todo, indagar acerca de cuánto sabían los estudiantes de este tema. Así se generó la encuesta acerca de “El ámbito del trabajo y tecnologías de la información”, que se suministró a los alumnos de la Cátedra que cursaron durante el primer cuatrimestre de 2015. Se analizaron 268 encuestas relevadas entre el 30 de abril y el 29 de mayo de 2015. Con respecto a la experiencia pedagógica se concluye que la respuesta que tuvo la convocatoria de participar en un trabajo de investigación fue mínima ya que implica un compromiso en el tiempo, más allá de la cursada de las materias, que no todos están en condiciones de asumir. Con respecto a la investigación cuantitativa algunos resultados son: el 78% de los encuestados fueron mujeres. El 90% considera al teletrabajo como mecanismo de inclusión social y laboral para personas discapacitadas y el 75%, un beneficio. Se propone fomentar la difusión del teletrabajo y seguir en este camino de la experiencia pedagógica en la formación de investigación, ya que la misma propicia la reflexión, la innovación y producción en los alumnos de la carrera.

Palabras claves: enseñanza; estadística; gestión; conocimiento; investigación; pedagógica

HACIA UN APRENDIZAJE PROFUNDO EN LA ASIGNATURA PROBABILIDAD Y ESTADÍSTICA NIVELES ALCANZADOS POR LOS ESTUDIANTES EN UNA ACTIVIDAD EVALUATIVA

NOEMÍ MARÍA FERRERI¹, RAÚL DAVID KATZ¹ y MARÍA EVANGELINA ÁLVAREZ²

¹FACULTAD DE CIENCIAS EXACTAS, INGENIERÍA Y AGRIMENSURA, UNR. ²FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA, UNR.

n.ferreri@fceia.unr.edu.ar

En la asignatura Probabilidad y Estadística, venimos trabajando para mejorar los procesos de enseñanza

y de aprendizaje, ideando permanentemente innovaciones en nuestra práctica docente. Es así que impulsamos en el aula una mayor participación de los estudiantes a través de la realización de actividades en el marco de un trabajo colaborativo, propiciando el planteo de preguntas, la confrontación de diferentes suposiciones y predicciones y la reestructuración de nociones erróneas. Esta forma de trabajo nos ofrece a los docentes interesantes elementos para actuar didácticamente, en beneficio de un mejor aprendizaje. Desde nuestro acompañamiento a los estudiantes en el desarrollo de conocimientos, habilidades, actitudes, etc. y desde las distintas instancias evaluativas, emerge la necesidad de innovaciones en nuestras prácticas, que posibiliten a los estudiantes reflexionar sobre su propio desempeño y tener conciencia sobre el alcance de la comprensión lograda. En este trabajo mostramos una actividad propuesta con el objetivo de indagar sobre la profundidad del aprendizaje alcanzado, en relación al concepto de variancia de un conjunto finito de datos. La misma involucró la definición, el cálculo, la interpretación y el reconocimiento de la misma a partir del gráfico de una distribución de frecuencias. Las respuestas dadas por los estudiantes revelan distintos niveles de aprendizajes alcanzados. Algunos testimonian un aprendizaje profundo por cuanto no sólo escriben correctamente la definición y la calculan correctamente, sino que además hacen una interpretación exhaustiva de la misma caracterizándola como un valor que cuantifica la variación de los datos respecto de la media. Más aún, hacen referencia al promedio de las desviaciones al cuadrado de los datos respecto a su media y logran caracterizarla a partir de un histograma. En contraposición encontramos estudiantes que no escriben la definición o lo hacen con errores, determinan a la variancia pero sólo dicen que la misma caracteriza la variación de los datos, sin especificar respecto de quién. Así mismo, para un histograma que sugiere una distribución uniforme de los valores, sostienen que la variancia es muy pequeña, por cuanto observan que las frecuencias relativas de los diferentes intervalos de clase presentan escasa variación entre ellas. El diálogo con los estudiantes, especialmente con los de menor desempeño, nos permitió comprender mejor el porqué de sus dificultades. Al finalizar esta actividad, nos encontramos ante nuevos desafíos: promover procesos que le permitan a los estudiantes reflexionar sobre sus propios saberes, formular nuevos interrogantes e integrar los conceptos a su estructura cognitiva.

Palabras claves: probabilidad y estad; evaluación; variancia; niveles de aprendiza

LA EXTENSIÓN DE LA ENCUESTA PERMANENTE DE HOGARES AL ÁMBITO RURAL: DESAFÍOS PARA LA FORMULACIÓN DE LA ESTRATEGIA DE MEDICIÓN URBANO-RURAL

CYNTHIA POK¹, MARISA DUARTE², SOLEDAD TRIANO² y DALILA VADELL³

¹INDEC. ²INDEC/UBA. ³INDEC/UBA/UNR.

dvadell@fcecon.unr.edu.ar

“La extensión de la Encuesta Permanente de Hogares al ámbito rural: desafíos para la formulación de la estrategia de medición urbano-rural” Cynthia Pok, Marisa Duarte, Soledad Triano, Dalila Vadell (INDEC/UBA/UNR) dvadell@fcecon.unr.edu.ar Las Encuestas a Hogares que llevan a cabo los Institutos Oficiales de Estadística, forman parte de las investigaciones orientadas al conocimiento de la realidad económico-social, generando indicadores intercensales que permiten caracterizar la situación social de los individuos y de los hogares, en un todo de acuerdo con las recomendaciones de los organismos internacionales y garantizando su comparabilidad. En nuestro país, la Encuesta Permanente de Hogares, programa de relevamiento continuo del Instituto Nacional de Estadística y Censos (INDEC), comenzó a relevar información sobre los hogares de Capital Federal en 1973, contemplando desde sus orígenes la extensión al total del país a través de sucesivas etapas. Desde el

punto de vista temático se diferenciaron dos etapas claramente definidas, la primera supuso la elaboración de un modelo adecuado al tratamiento de indicadores sociales en el contexto de mercados de trabajo urbanos de carácter permanente y la segunda al ámbito de los mercados de trabajo urbano-rurales de carácter estacional. En su extensión espacial se fueron incorporando progresivamente aglomerados urbanos de más de 100.000 habitantes y todas las capitales provinciales, al tiempo que se iniciaba el tratamiento de indicadores sociales en el contexto de los mercados de trabajo urbano-rurales de carácter no permanente, a partir de dos experiencias: Alto Valle del Río Negro y provincia de Tucumán. En la actualidad se cubre como dominios de estimación 31 aglomerados y el total urbano en el tercer trimestre de cada año. La etapa de extensión iniciada en 2016, pretende avanzar en la caracterización, medición y análisis del mercado de trabajo para el conjunto de la población del país. El presente trabajo aborda los principales nudos conceptuales que plantea el entorno de los mercados laborales afectados por la estacionalidad, muestra los avances en términos de pruebas temáticas en curso, ahondando sobre las características que requieren ser visibilizadas en cuanto a las estrategias de reproducción y de inserción laboral propia de los contextos semiurbanos y rurales (trabajo de producción para el autoconsumo, antigüedad cíclica, ingresos no monetarios entre otros aspectos), expone los resultados iniciales de dichas pruebas y los desafíos a encarar para la formulación de la estrategia de medición urbano-rural.

Palabras claves: estadísticas oficial; encuestas a hogares; ámbito rural

LA POBREZA E INDIGENCIA MULTICAUSAL: UNA PROPUESTA METODOLÓGICA PARA LA PROVINCIA DE MISIONES

SILVANA DEA LABAT y DARIO EZEQUIEL DÍAZ

INSTITUTO PROVINCIAL DE ESTADÍSTICA Y CENOS DE LA PROVINCIA DE MISIONES.

slabat24@hotmail.com

Resumen En los últimos años, la pobreza e indigencia se han vuelto un fenómeno complejo para lograr una medición confiable, válida y lo más objetiva posible. En la literatura académica existe una multiplicidad de enfoques y abordajes con el objetivo concreto de pretender comprender lo que implica ser pobre, para evitar la subestimación en el recuento de los mismos. Recientemente, la “multidimensionalidad” de la pobreza ha alcanzado un profundo reconocimiento, pues se divisa una inquietud en adoptar metodologías con posibilidad de diferenciar tópicos de índole social o nuevas dimensiones humanas. Generalmente, en la literatura, la pobreza se ha vinculado de manera habitual con los conceptos de necesidad estándar de vida, o insuficiencia de recursos, pues se trata de atributos sobre los que pueden obtenerse medidas relativamente aceptables. En tanto, el marco teórico que se utilice y su correspondiente metodología están estrechamente relacionados con los objetivos de la investigación que se esté llevando a cabo y sus resultados pueden contribuir a identificar aspectos particulares de la pobreza. Aun cuando existe una gran pluralidad de aproximaciones teóricas para reconocer qué hace pobre a un individuo, hay un acuerdo cada vez más amplio sobre la naturaleza multidimensional de este concepto, el cual afirma que los elementos que toda persona precisa para tomar una decisión de manera libre, informada y con igualdad de oportunidades sobre sus alternativas vitales, no pueden ser reducidos a una sola de las características o dimensiones de su existencia. Los abordajes alternativos para medir la pobreza, el nivel de vida y el desarrollo que parten del rechazo explícito a la posibilidad de hallar un patrón de medición única y universal, se transforman inevitablemente en enfoques multidimensionales. El presente trabajo tiene como objetivo primordial presentar la metodología utilizada por el Instituto de Estadística y Censos de la Provincia de Misiones para el diseño del cálculo del Índice de Pobreza Multicausal para la Provincia de Misiones. El Índice de Pobreza e Indigencia Multidimensional se refiere a una estructura metodológica que busca dar respuesta a un problema a partir de la identificación de un conjunto de

causas posibles que lo generan. El mismo surge del promedio ponderado de tres métodos: • Método Indirecto Absoluto (MIA) • Método Indirecto Relativo (MIR) • Método Directo de Satisfacción de Necesidades Básicas (MDS) La ponderación de esta tasa de pobreza e indigencia multicausal fue la siguiente: 25% de MIA, 25% MIR y 50% MDS.

Palabras claves: pobreza multicausal; medición; bienestar; ingresos

R-LADIES: MEJORANDO LA DIVERSIDAD DE GÉNERO EN LA COMUNIDAD R

NATALIA DA SILVA¹ y LAURA ACIÓN²

¹IESTA, UDELAR. ²UBA.

ndasilva@iastate.edu

R es un lenguaje de programación de código abierto, está desarrollado por la comunidad y es ampliamente utilizado en educación, investigación, y la aplicación de la estadística tanto en la academia como la industria. De acuerdo a resultados presentados en la conferencia Mujeres en Estadística y Ciencia de Datos 2016 (organizada por la Asociación Americana de Estadística), las mujeres y todas las personas que se identifican significativamente como mujer están subrepresentadas dentro de la comunidad R. Aproximadamente, sólo 17% de los paquetes de R son creados y/o mantenidos por mujeres. Además, de los integrantes ordinarios de la R Foundation, el organismo que apoya el desarrollo de R, sólo 13.5% son mujeres. Este último porcentaje superó la barrera del 10% sólo recientemente. Finalmente, el equipo de liderazgo de R cuenta sólo con una mujer entre sus 7 integrantes. R-Ladies surge como una iniciativa para revertir la subrepresentación de las mujeres en la comunidad R. El objetivo de R-Ladies es mejorar la participación femenina no sólo como usuarias, desarrolladoras, y oradoras en conferencias, sino también como líderes en la comunidad. R-Ladies es una organización internacional que incentiva la diversidad de género en la comunidad R a través de grupos de usuarios en distintas ciudades. R-Ladies organiza reuniones presenciales, actividades virtuales y mentorías en un entorno informal y amigable. Organizaciones similares en ciencias, tecnología, ingeniería y matemática muestran que este tipo de iniciativas ayudan a mejorar la representación de las minorías en estos ámbitos. El primer grupo de R-Ladies se organizó en Octubre 2012. En 2016 surgió la organización madre R-Ladies Global a partir de la cual R-Ladies se expandió a más de 45 grupos en ciudades de más de 20 países. R-Ladies cuenta actualmente con más de 6000 integrantes. Cada grupo reúne a mujeres y hombres tanto de la academia como de la industria. Los integrantes provienen de diferentes disciplinas, tienen diferentes grados de conocimiento de R y están en diferentes etapas de su carrera. En esta charla, presentaremos estadísticas de los grupos de R-Ladies en Uruguay y Argentina. También mostraremos estadísticas sobre el desarrollo de R-Ladies en América Latina y en el resto del mundo. Además, compartiremos nuestras experiencias relacionadas con la logística de la gestión de grupos locales. Presentaremos historias de éxito en todo el mundo, algunos de los desafíos que se enfrentan y cómo hacer para sumarse a un grupo existente o para crear uno nuevo.

Palabras claves: comunidad r; diversidad de género; software libre; talleres de r; brecha de género; mujeres liderando

RECONOCIMIENTO DE TEXTURAS EN UNA IMAGEN POLARIMÉTRICA BASADO EN DISTANCIAS ESTOCÁSTICAS

SILVINA PISTONES¹, JORGE MARTINEZ², ÉRICA SCHLAPS³ y ALEJANDRO FRERY⁴

¹UNIVERSIDAD NACIONAL DEL SUR Y TECNOLÓGICA NACIONAL REGIONAL BAHÍA BLANCA. ²UNIVERSIDAD NACIONAL DEL SUR.
³UNIVERSIDAD NACIONAL DE LA PAMPA. ⁴LABORATÓRIO DE COMPUTAÇÃO CIENTÍFICA E ANÁLISE NUMÉRICA. UNIVERSIDADE FEDERAL DE ALAGOAS.

lpistone@criba.edu.ar

Las imágenes de Radar de Apertura Sintética (Synthetic Aperture Radar — SAR) tanto monopolarizadas como polarimétricas, son fundamentales para el estudio y caracterización de la zona observada, dado que a partir de ellas puede obtenerse información que ningún otro tipo de sensor provee. Sin embargo, estas imágenes tienen la desventaja de que son difíciles de analizar e interpretar debido, entre otras razones, a la presencia de ruido speckle, que es inherente al proceso de captura y formación de imágenes con iluminación coherente. La extracción de características de textura, intensidad, posición espacial, color, granularidad, etc., de una imagen constituye una etapa previa para el procesamiento automático de imágenes relacionado con la segmentación, la clasificación, y la detección e identificación de objetos. Este trabajo pretende evaluar la capacidad de las distancias estocásticas en el reconocimiento de diferentes texturas (distintos tipos de cobertura terrestre) presentes en una imagen polarimétrica. Para el estudio, se selecciona una subimagen de 100x512 píxeles, de la imagen polarimétrica de 4-looks de la Bahía de San Francisco, que presenta tres texturas bien diferenciables correspondientes a bosque, zona urbana, y agua. Inicialmente se efectúa un análisis cuantitativo descriptivo de estos tres conjuntos de datos de diferentes texturas. Hecho esto, suponemos que los datos, que son de polarización completa, obedecen distribuciones Wishart Complejas. Con el fin de discriminar las regiones de interés, de cada una de ellas se seleccionan dos muestras de tamaño 20x20 píxeles y se calculan las distancias estocásticas de Kullback-Leibler (d_{KL}), Bhattacharyya (dB) y la de Hellinger (dH), intra y entre los conjuntos de datos. Para ello fue necesario estimar previamente los parámetros de las distribuciones Wishart de dichos conjuntos de datos mediante el método de Máxima Verosimilitud. El análisis descriptivo de las distintas texturas refleja una intensa asimetría a derecha y una gran presencia de datos que no se explican bien por una ley gaussiana (outliers). En todas las comparaciones de texturas consideradas, los valores de las distancias respetaron el siguiente orden: $d_{KL} \leq dB \leq dH$. Las distancias d_{KL} y dB permiten discriminar con éxito las diferentes, alcanzando valores menores entre áreas cuyo grado de rugosidad es relativamente similar. En cambio, la distancia dH evidencia ausencia de poder discriminativo de las diferentes rugosidades de las texturas bajo estudio.

Palabras claves: imagen polarimétrica; texturas; distancias estocásticas; estimador de máxima

UNA PROPUESTA DE EVALUACIÓN DE CONCEPTOS ESTADÍSTICOS EN CARRERAS DEL ÁREA SOCIALES

SILVANA SANTELLÁN, LILIANA TAUBER y MARIELA CRAVERO

UNIVERSIDAD NACIONAL DEL LITORAL.

estadisticamatematicafhuc@gmail.com

Compartimos una propuesta de evaluación continua que pretende enriquecer las prácticas de enseñanza de Estadística para Ciencias Sociales, teniendo como premisa desarrollar el pensamiento estadístico crítico. Nuestra Propuesta Uno de los principales desafíos a los que nos enfrentamos los educadores

estadísticos, en carreras sociales, es proponer una enseñanza que logre desarrollar aprendizajes a largo plazo. En este sentido, coincidimos con Behar y Pere Grima (2014), Herrera y Konic (2017), quienes proponen algunos lineamientos para introducir contenidos de Estadística que sean significativos para los estudiantes. Este desafío es el que nos hemos propuesto desde la cátedra “Métodos estadísticos para Ciencias Sociales”, en la Facultad de Humanidades y Ciencias de la Universidad Nacional del Litoral. Desde el inicio de este espacio observamos que los alumnos comienzan el cursado indicando que aprender los contenidos estadísticos es tarea difícil y, una baja proporción de alumnos lograba realizar exitosamente el recorrido. Si bien estos alumnos evidencian mucha autonomía en la lectura de textos específicos de las Ciencias Sociales, no la tienen cuando deben leer textos de carácter más técnico como puede ser uno de Estadística. Además, los estudiantes no acostumbran a cursar regularmente, lo cual es un problema agregado para la planificación de las clases, ya que una gran proporción abandona rápidamente. Este contexto nos llevó a diseñar una propuesta que priorice la participación activa de nuestros estudiantes y cuyo eje mentor sea la continuidad de los contenidos desarrollados. Delineamos una evaluación continua atravesada por tres trabajos prácticos basados en un texto de Escudero (2014), el software Gapminder y los últimos datos disponibles de la Encuesta Permanente de Hogares (EPH) del aglomerado Gran Santa Fe. El objetivo de estos prácticos es promover el razonamiento crítico trabajando las ideas estadísticas fundamentales que plantea Batanero (2013), para fomentar la formación de profesionales cultos estadísticamente. Reflexiones Finales Consideramos que nuestros estudiantes han tenido la posibilidad de lograr aprendizajes más significativos de los contenidos estadísticos; principalmente debido al importante rol del análisis sobre situaciones reales y el reconocimiento de determinados conceptos estadísticos en estas situaciones y la interpretación de medidas obtenidas sobre estos datos.

Palabras claves: cultura estadística; razonamiento estadístico; evaluación continua; estudiantes universitarios

UNA PROPUESTA DIDÁCTICA QUE PROPICIA EL TRABAJO CON IDEAS ESTOCÁSTICAS FUNDAMENTALES Y EL RAZONAMIENTO INFERENCIAL INFORMAL

YANINA REDONDO, LARISA ZILLONI y LILIANA TAUBER

UNIVERSIDAD NACIONAL DEL LITORAL.

estadisticamatematicafhuc@gmail.com

A partir de trabajos realizados en el marco del proyecto CAI+D: “La inferencia informal como objetivo central de la educación estadística en estudiantes universitarios”, hemos identificado una serie de conceptos asociados a ideas fundamentales del razonamiento informal y de la inferencia estadística informal, estableciendo categorizaciones de dichos conceptos de acuerdo al grado de influencia sobre la comprensión de estas ideas fundamentales (Bianchi y Tauber, 2014; Gioria, 2015). Por otra parte, a través de experiencias previas (Tauber, 2014), hemos detectado un escaso nivel en la formación estadística de los profesores de matemática. Esto evidencia un grave problema, fundamentalmente si consideramos que la educación estadística puede brindar elementos importantes para el pensamiento crítico de cualquier ciudadano. A partir de la identificación de los elementos principales de nuestro marco teórico y del análisis previo de la situación, hemos podido delimitar los dos objetivos que perseguimos en este trabajo: Presentar una propuesta de enseñanza para el aula secundaria que permita construir significados acerca de las ideas fundamentales de aleatoriedad, variabilidad, distribución y modelo, y Realizar un análisis de contenido basado en la caracterización de razonamientos informales que puedan observarse al desarrollar una actividad que integra conceptos del Análisis Exploratorio de Datos y la Inferencia Estadística Informal. Como las actividades

propuestas aún no han sido implementadas, aún no podemos hablar de resultados en ese sentido. Pero sí podemos indicar que, dado que el presente forma parte de las tareas programadas en el plan de trabajo de una adscripción en investigación desarrollada por una de las autoras, el resultado de este trabajo es la elaboración de la actividad que presentamos. Por otra parte, en base al análisis de contenido hemos podido establecer una red conceptual en la que se describen las relaciones entre los conceptos e ideas fundamentales que se ponen en juego a partir del desarrollo y ejecución de la propuesta didáctica y, también hemos podido establecer una categorización de los tipos de razonamientos estocásticos informales que se podrían poner en correspondencia A modo de conclusión, consideramos que hemos aportado una propuesta que tiene un enorme potencial didáctico ya que permite poner en relación diversas ideas estocásticas fundamentales y tipos de razonamientos que pueden propiciar la generación de los fundamentos de la Inferencia Estadística Informal.

Palabras claves: ideas estocásticas f; razonamiento estadís; inferencia estadísti

USO DE INDICADORES PARA EL ANÁLISIS DE PROBLEMAS PROPUESTOS EN TEXTOS DE PROBABILIDAD Y ESTADÍSTICA PARA INGENIERÍA

NOEMÍ MARÍA FERRERI¹, GRACIELA HAYDÉE CARNEVALI¹ y MARÍA EVANGELINA ÁLVAREZ²

¹FACULTAD DE CIENCIAS EXACTAS, INGENIERÍA Y AGRIMENSURA, UNR. ²FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA, UNR.

nferreri@fceia.unr.edu.ar

La resolución de problemas de naturaleza estadística, es una competencia importante a desarrollar en los futuros ingenieros. Esta tarea implica un conjunto de procesos que se articulan formando un ciclo, al que, por ejemplo, Wild y Pfannkuch, denominan “Ciclo PPDAC”. La sigla corresponde a las etapas “planteo del problema” (P), “planificación del estudio estadístico” (P), “recolección de los datos” (D), “análisis de los mismos” (A) y “obtención de conclusiones”, en contexto (C). Con el objetivo de que los futuros ingenieros adquieran esta competencia, es deseable que en los cursos de Estadística se resuelvan frecuentemente problemas diseñados de tal manera que, en el proceso de resolución, los alumnos deban transitar la mayor cantidad de etapas y se enfrenten a diferentes situaciones. Para facilitar el diseño de problemas que tengan estas características, se cuenta con un conjunto de indicadores propuestos en trabajos anteriores, el cual se viene elaborando en el marco del proyecto de investigación ING525 “El Pensamiento Estadístico en el control y la mejora de los procesos: diseño, aplicación y evaluación de propuestas didácticas para su desarrollo en alumnos de Ingeniería Industrial”. En muchas ocasiones los problemas se toman de textos de referencia, que no siempre contienen todos los elementos necesarios para las diferentes etapas del ciclo PPDAC. El uso de los indicadores facilita el análisis de estos problemas y orienta a los docentes para que los enriquezcan, favoreciendo así el desarrollo de la competencia deseada. En el presente trabajo se analizan algunos problemas tomados de textos de Probabilidad y Estadística de uso habitual en carreras de Ingeniería, aplicando los indicadores mencionados y, a modo de ejemplo, se hacen algunas propuestas para enriquecerlos.

Palabras claves: resolución de proble; ciclo ppdac; indicadores; textos de probabilidad

PÓSTERS

Datos categóricos

DETERMINANTES DE LA INNOVACION UN ANÁLISIS A NIVEL DE FIRMA POR TAMAÑO Y SECTOR

VERÓNICA ARIAS, LAURA ISABEL LUNA, MARÍA INÉS AHUMADA y NORMA PATRICIA CARO

FACULTAD DE CIENCIAS ECONÓMICAS, UNIVERSIDAD NACIONAL DE CÓRDOBA..

veroarias1@gmail.com

A nivel de empresa, la innovación conduce a un uso más eficiente de los recursos, generando así ventajas competitivas sostenibles en el tiempo. Se reconoce a la innovación como uno de los factores claves de la competitividad y el crecimiento, por lo cual existe gran interés en comprender las variables que la afectan. La “Encuesta sobre innovación y conducta tecnológica de la Provincia de Córdoba” llevada a cabo por la Dirección General de Estadística y Censos, es fuente de información y los datos que provee son las observaciones y mediciones presentadas por cada empresa a través del tiempo, logrando conformar un panel de datos para este estudio. Estos cuentan con una estructura jerárquica (empresas medidas repetidamente en el tiempo), la que introduce dependencia en las respuestas múltiples dentro de cada unidad. El presente trabajo tiene como objetivo generar un aporte a la comprensión del proceso de innovación en las empresas industriales manufactureras de la provincia de Córdoba para el período 2011 a 2014, incorporando nuevas dimensiones de análisis a los resultados obtenidos en trabajos previos sobre los principales determinantes de la innovación en las empresas. En esta oportunidad de análisis se introducen variables de clasificación de las empresas según el sector al que pertenecen y el tamaño de las mismas. En el primer caso se considera la clasificación realizada por agrupación sectorial a saber: intensivas en I+D, escala, trabajo y en recursos naturales. Para la clasificación por tamaño se realizó una división entre empresas grandes y PyMes. Luego, se buscaron los principales determinantes de la innovación ajustando modelos logísticos mixtos, donde la variable respuesta considera si la empresa logró nuevos productos y/o procesos o bien mejoras significativas. Mientras que los indicadores vinculados a la estructura de las firmas, factores sectoriales y determinantes sistémicos se asumen como variables predictoras. Resultaron significativos en el estado innovador de la empresa, dependiendo del sector y tamaño considerado, la proporción de profesionales ocupados en la empresa, los gastos efectuados en actividades de innovación interna y externa, exportar, la utilización de fondos de programas oficiales de innovación y la vinculación con organismos públicos del sistema nacional de innovación y el tamaño.

Palabras claves: innovación; modelo logístico mix; industria manufactur; provincia de córdoba

HETEROGENEIDAD DE LAS PREFERENCIAS DE LOS CONSUMIDORES DE ALIMENTOS DE CALIDAD DIFERENCIADA

BEATRIZ LUPÍN

GRUPO ECONOMÍA AGRARIA, FCEYS-UNMDP.

beatrizlupin@gmail.com

El objetivo es estudiar la heterogeneidad en las preferencias de los consumidores. A tal fin, se desarrollan aspectos conceptuales de los Modelos Logit Condicional (MLC) y Logit Mixto (MLM),

presentando resultados obtenidos en la primera etapa de la investigación. Mediante los experimentos de elección (EE), los consumidores declaran sus preferencias, seleccionando, por bloques, alternativas de un bien -combinación de niveles de atributos-. Se basan en el Modelo de Utilidad Aleatoria (MUT): $Unik = B' Xnik + Enik$ - $Unik$ = utilidad proporcionada por la alternativa i del bloque k al individuo n ; B = vector de coeficientes no observados; $Xnik$ = vector de variables observadas; $Enik$ = término de error, iid Valor-Extremo Tipo I (Gumbel)-. El MLC estima un parámetro común para cada atributo. La heterogeneidad es incorporada realizando estimaciones por segmentos de consumidores según variables socio-económicas o a través de la interacción de los atributos con dichas variables. Por su parte, el MLM captura directamente la heterogeneidad, presentando dos versiones, conforme el tratamiento dado al MUT: Parámetros Aleatorios: centrada en las variaciones de las preferencias: $Unik = Bn' Xnik + Enik$ - Bn = vector de coeficientes no observados para cada individuo, varían en la población-. Componentes del Error: centrada en los patrones de sustitución de las alternativas: $Unik = B' Xnik + (Nnik + Enik)$ - $Nnik$ = término aleatorio, correlacionado sobre las alternativas-. En ambos casos, mediante simulación, se aproxima la probabilidad de elección, asumiendo una determinada distribución para los coeficientes o para N , según corresponda. La selección de la distribución constituye un reto empírico. Con datos de un EE referido a papa producida con bajo impacto ambiental (Mar del Plata, octubre 2012, 402 casos), se aplicó un MLC de efectos principales, segmentando la muestra por nivel socio-económico. Luego, con los coeficientes estimados, se calculó la disposición a pagar por los atributos "contenido de agroquímicos" y "aptitud culinaria". Los participantes de nivel socio-económico medio (45% de la muestra) son los están dispuestos a pagar más: 4 y 2, adicionales, por papa con bajo contenido de agroquímicos y de buena calidad culinaria, respectivamente, de lo que pagan por papa sin dichas cualidades. Estas especificaciones describen limitadamente la heterogeneidad pues no indican la variación en las preferencias de individuos con iguales características. Dado que una adecuada identificación de la heterogeneidad mejora la potencia explicativa, se tiene previsto continuar la investigación con la aplicación del MLM.

Palabras claves: alimentos; consumidores; calidad diferenciada; preferencias; heterogeneidad

MODELO LOG-ACUMULADO MIXTO APLICADO A LA TASA RESPIRATORIA BAJO ESTRES SEVERO DEBIDO A TEMPERATURAS EXTREMAS EN OVINOS

NATALIA RUBIO¹, M FERNANDA LOPEZ ARMENGOL² y GUSTAVO GIMÉNEZ¹

¹DEPARTAMENTO DE ESTADÍSTICA, FACULTAD DE ECONOMÍA Y ADMINISTRACIÓN UNCO. ²CENTRO DE INVESTIGACIONES EN TOXICOLOGÍA AMBIENTAL Y AGROBIOTECNOLOGÍA DEL COMAHUE.

rubionatalia@yahoo.com.ar

El clima es uno de los factores más importantes a tener en cuenta en la cría de animales debido a que los cambios en las variables climáticas como temperatura, humedad, movimiento del aire, fotoperiodo y radiación solar no pueden ser controlados, pero si estudiados. Los mismos afectan de forma directa el aspecto productivo y reproductivo de ovejas. Se ha estudiado que las mismas cuentan con una gran habilidad para adaptarse a distintas condiciones climáticas. La adaptación y habituación han hecho y siguen haciendo posible la vida en los distintos ambientes. En relación con ello es notable que la oveja se considere un animal con un espectro extremadamente amplio de adaptación a los diversos climas con diferentes condiciones micro ambientales y de disponibilidad de alimentos; destacándose su alta capacidad de adaptación en ambientes donde prevalecen climas fríos o templados, áridos o secos y tropicales y húmedos. Sin embargo, a pesar de esta versatilidad fisiológica y biológica de tolerar diversas condiciones ambientales, el estrés térmico repercute negativamente en los procesos vitales y funciones relacionadas con la reproducción animal. Se desarrolló

una experiencia en la localidad de Cinco Saltos, Río Negro, que cuenta con condiciones climáticas controladas en una cámara de calor, resultando éste estudio de relevancia regional. Se midió la tasa respiratoria, número de jadeos por minuto, a seis carneros en condiciones físicas similares a diferentes temperaturas para evaluar el estrés generado por el calor. Para analizar el efecto que tiene la lana en los animales, tres de ellos fueron esquilados y tres se dejaron sin esquilar. También se registró la posición del animal durante el estudio, debido a que puede afectar su tasa respiratoria. Se consideró la respuesta según cinco categorías ordinales que permitió construir un modelo log-acumulado mixto, que contempla la variabilidad entre carneros, posición del carnero (parado – sentado), condición (esquilado – no esquilado) e índice de temperatura-humedad (ITH). Se observó un efecto significativo de los individuos. En la estimación de los parámetros del modelo se observa una interacción triple significativa. Se continuó analizando dentro de la condición de los carneros. Para los esquilados a medida que se incrementa el ITH, la chance de aumentar de categoría respiratoria es 2 veces mayor; mientras que, para los no esquilados es 1,5 veces. En ambos casos no importa la posición en que se encuentra el carnero.

Palabras claves: estrés térmico; respuesta ordinal; modelo multicategóri

Datos de duración

EXTENSIÓN DEL MENÚ EPIDEMIOLOGÍA DEL SOFTWARE INFOSTAT: PROCEDIMIENTOS PARA ANÁLISIS DE SUPERVIVENCIA

MARGOT TABLADA¹, LAURA ALICIA GONZALEZ¹, PABLO PACCIORETTI¹,
DUBERNEY MUÑOZ MORALES² y DAVI MORALES¹

¹FCA. ²UNC.

gonzalez.lauraalicia@gmail.com

EXTENSIÓN DEL MENÚ EPIDEMIOLOGÍA DEL SOFTWARE INFOSTAT: PROCEDIMIENTOS PARA ANÁLISIS DE SUPERVIVENCIA TABLADA ELENA MARGOT¹, GONZALEZ LAURA ALICIA¹, PACCIORETTI PABLO², MUÑOZ MORALES DUBERNEY³, MORALES DAVID¹, 1Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba. 2Estudiante de Maestría en Estadística Aplicada, Universidad Nacional de Córdoba. 3Estudiante de doctorado en Ciencias de la Ingeniería, Universidad Nacional de Córdoba mtablada@gmail.com RESUMEN El seguimiento de individuos para conocer el tiempo transcurrido hasta que se presenta un resultado de interés es una práctica en numerosas áreas como agronomía, ecología, medicina, psicología, economía, sociología, industria, entre otras. Los métodos para el tratamiento estadístico de los datos se conocen como análisis de supervivencia o de sobrevida. El evento de interés no necesariamente es la muerte y, por ejemplo, se puede estudiar el tiempo hasta que falla un producto o equipo, o el tiempo hasta la reincidencia en el delito, entre otros. Es frecuente que los registros de los tiempos estén acompañados de otras variables (edad, material, sexo, tratamiento, ingresos, estado civil, etc.), que se usan como clasificadorias de grupos o como covariables relacionadas con el evento de interés. Mediante el análisis se pretende estimar la función de supervivencia y la función de riesgo, y llevar a cabo comparaciones entre grupos. El módulo de aplicaciones en Epidemiología, del software estadístico InfoStat posee procedimientos para ajuste de tasas, obtención de índices de concordancia y consistencia, medidas de riesgos y asociación para diferentes tipos de muestreo, curva ROC y regresión de Cox. Mediante la extensión propuesta, se incorporan a éstos técnicas tradicionalmente utilizadas para la descripción, estimación y contraste de hipótesis en el análisis

de sobrevida: estimación de la función de supervivencia (Kaplan-Meier, Flemming y Harrington), medidas resumen, intervalos de confianza, gráficos, comparación de curvas (prueba log-rank, de Wilcoxon, Tarone, Peto). También se incluyen rutinas que amplían los resultados obtenidos con el uso del módulo actual para la regresión de Cox: función de supervivencia ajustada mediante modelo de Cox, herramientas para verificar supuestos (riesgos proporcionales, residuos tipo score, deviance, martingala). La implementación se complementa con la elaboración de bases de datos y un tutorial, para ejemplificar el uso de los procedimientos, que incluye un marco conceptual introductorio. El desarrollo se realiza mediante la combinación de herramientas de programación de InfoStat y R que permite ofrecer aplicaciones, en base a paquetes internacionalmente conocidos y validados, en una interfaz amigable. El desarrollo del módulo Epidemiología es resultado de proyectos avalados por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de Córdoba.

Palabras claves: bioestadística; estudios de seguimiento; software estadístico

SEGUIMIENTO DE PACIENTES TRATADOS POR UNA INFECCIÓN PARASITARIA CRÓNICA: REVISIÓN SISTEMÁTICA Y METAANÁLISIS DE PARTICIPANTES INDIVIDUALES

KAREN N. ROBERTS¹, GUILLERMINA B. HARVEY², CRISTINA B. CUESTA²,
YANINA SGUASSERO¹, AGUSTÍN CIAPPONI³ y SERGIO SOSA-ESTANI⁴

¹CENTRO ROSARINO DE ESTUDIOS PERINATALES. ²FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA, UNR. ³INSTITUTO DE EFECTIVIDAD CLÍNICA Y SANITARIA. ⁴INSTITUTO NACIONAL DE PARASITOLOGÍA FATALA CHABEN-ANLIS / CONSEJO NACIONAL DE INVESTIGACIONES CIENTÍFICAS Y TÉCNICAS.

robertskaren@gmail.com

La enfermedad de Chagas es una infección endémica de América Latina causada por el parásito *Trypanosoma cruzi* (*T. cruzi*). En la fase crónica de la enfermedad, el éxito del tratamiento es incierto porque la seronegativización (desaparición de anticuerpos contra *T. cruzi*) puede tardar décadas. Con el propósito de conocer más acerca del tiempo hasta la seronegativización, se realizó una revisión sistemática con meta-análisis de datos de participantes individuales (Nº PROSPERO: CRD42012002162). Se incluyeron estudios de cohortes e ICAs. Las bases de datos originales de los pacientes individuales fueron solicitadas a los investigadores principales de cada estudio. Los resultados primarios se dicotomizaron como positivos o negativos para las pruebas de inmunoadsorción enzimática (ELISA), inmunofluorescencia indirecta (IIF) y análisis de hemaglutinación indirecta (IHA). Se estimaron las funciones de supervivencia y de riesgo aplicando el método de Kaplan-Meier, y se ajustaron modelos de regresión de riesgos proporcionales de Cox. A los fines de considerar la variabilidad entre las distintas poblaciones, se consideró un efecto aleatorio asociado a cada estudio. El exceso de riesgo o fragilidad para distintos estudios se describió utilizando el componente de variancia del efecto aleatorio y el correspondiente coeficiente Tau de Kendall. Los análisis de subgrupos se basaron en edad al tratamiento (menor vs mayor de 19 años) y región donde se adquirió la infección (Argentina, Bolivia, Chile, Paraguay con predominio de cepa TcV vs Brasil con predominio de cepa TcII). Los objetivos del trabajo fueron determinar, mediante la aplicación de métodos estadísticos apropiados, la evolución de las pruebas serológicas en sujetos con infección crónica por *T. cruzi* tratados con fármacos tripanocidas y describir el efecto de los factores explicativos sobre la seronegativización. Se incluyeron 27 estudios y se obtuvieron datos de 1.296 sujetos. El riesgo de sesgo fue bajo en 17 estudios (63%). El valor del coeficiente Tau de Kendall fue de 0.63, 0.58 y 0.47 para ELISA, IIF y IHA, respectivamente. Basados en el conjunto de datos disponible, los hallazgos muestran una interacción entre la edad al tratamiento y la región. En la región TcII la posibilidad de seronegativización fue mayor en los sujetos tratados entre 1-19 años comparados con adultos; HR 6.60 (IC 95% 2.03-21.42) para ELISA, HR 9.37 (3.44-25.50) para IIF y HR 5.55 (1.46-21.11) para IHA.

Palabras claves: metaanálisis de dato; tasa de supervivencia; análisis de regresión; trypanosoma cruzi; enfermedad crónica

Datos faltantes

CARACTERIZACIÓN CLIMATOLÓGICA PARA UNA REGIÓN DE LA PROVINCIA DE SANTA CRUZ DESPROVISTA DE INFORMACIÓN

JULIO SOTO¹, DORA SILVIA MAGLIONE¹ y OSCAR BONFIL²

¹UNIVERSIDAD NACIONAL DE LA PATAGONIA AUSTRAL. ²SERVICIO METEOROLÓGICO NACIONAL, OFICINA METEOROLÓGICA RÍO GALLEGOS.

dmaglione@disytel.net

Muchas actividades económicas están supeditadas a las condiciones climáticas de la región en la que se desarrollan ya que todo ser vivo depende de las condiciones de la temperie. Sin contar con datos meteorológicos representativos de una zona, difícilmente se puedan planificar adecuadamente actividades de toda índole. La provincia de Santa Cruz, y en particular la región central es una de las zonas más desprovista de información meteorológica de largo plazo (climatológica) a nivel mundial, y a su vez es una de las áreas mineras más importante de la Patagonia. Con la finalidad de poder caracterizar un sector de la región (de la cual no se poseen registros) se trabajó con información proveniente de distintas fuentes tales como el Servicio Meteorológico Nacional, la Administración General de Vialidad Provincial, el Instituto Nacional de Tecnología Agropecuaria, el Instituto Nacional de Investigación Agropecuaria de Chile, entre otros. Algunas fenómenos discontinuos como las lluvias tienen una variabilidad temporal mayor por lo que se puede llegar a necesitar más datos. En este trabajo se utiliza la metodología usada por Cressman (1959) para lograr una aproximación iterativa para la elaboración de la variable objetivo en cada punto de una región usando como funciones de pesos las distancias inversas de las redes de observaciones que se poseen. A partir de este método se analizan y caracteriza de forma aproximada a las siguientes variables meteorológicas en forma anual y estacional: Temperatura Media; Temperatura Máxima; Temperatura Mínima; Precipitación; Humedad Relativa; Presión Atmosférica y Velocidad Media de Viento en una región de la provincia de Santa Cruz de la que no existen registros meteorológicos. Además se analizó el Índice de Concentración de Precipitación (ICP); Índice de Aridez De Demarthonne y la Evapotranspiración de la FAO (ETo – FAO).

Palabras claves: variables climatológicas; método de la distancia; mapas

COMPARACIÓN DE MÉTODOS PARA COMPLETAR DATOS FALTANTES EN VARIABLES CLIMATOLÓGICAS EN PATAGONIA SUR

DORA SILVIA MAGLIONE¹, JOSÉ LUIS SÁENZ¹, JULIO SOTO¹, OSCAR BONFILI², CARLOS TALAY¹, MARISA SANDOVAL¹ y MIGUEL LLANCALAHUEN¹

¹UNIVERSIDAD NACIONAL DE LA PATAGONIA AUSTRAL. ²SERVICIO METEOROLÓGICO NACIONAL, OFICINA METEOROLÓGICA RÍO GALLEGOS.

dmaglione@disytel.net

La región Patagónica sur se caracteriza por su amplia extensión y por la escasez de estaciones meteorológicas en donde recabar información. En este trabajo se utilizan los datos de estaciones meteorológicas del INTA de la provincia de Santa Cruz y sur de Chubut (todas distantes y con presencia de datos faltantes), para comparar seis métodos propuestos en la literatura para completar series mensuales, como son el método de Karl, el método CLP, el método de la razón q, el método de la razón normal, la regresión lineal simple, la correlación múltiple y el método de las distancias inversas para las variables temperatura media, máxima y mínima, humedad relativa, precipitación acumulada y presión atmosférica. Para poder llevar adelante las comparaciones de estos métodos se realiza un estudio sobre los errores obtenidos al predecir los datos de valores existentes al ser eliminados de la serie, usando distintas medidas de selección y de interpretación en cada una de las estaciones. Para las cuatro primeras variables, la correlación es alta entre todas las estaciones climatológicas, por lo que son consideradas todas ellas al aplicar los siete métodos; se observa que los mejores métodos para completar datos faltantes resultan ser el de regresión lineal, el de correlación múltiple y el de la distancia inversa, con variaciones según la variable y las ubicaciones geográficas de las estaciones meteorológicas. En el caso de las variables presión atmosférica y precipitación acumulada mensual, no existe correlación alta entre todas las estaciones, por lo que los siete métodos se aplican sólo a grupos más reducidos que presentan correlación moderada entre las series mensuales; en este caso se observa que los mejores métodos para completar las series de presión atmosférica resultan ser el de regresión lineal, el de correlación múltiple y el de la distancia inversa, y para la precipitación acumulada aparece mayor diversidad de métodos dependiendo de la ubicación geográfica como son la razón q, regresión lineal, distancia inversa, razón normal y el método CLP. Para estas dos últimas variables si se utilizan métodos que no tengan en cuenta las correlaciones entre estaciones sino toda la información disponible en la región, se obtiene que el método de la distancia inversa es el que mejor completa la información faltante para la precipitación acumulada mensual en todas las estaciones climatológicas, en tanto que para la presión atmosférica principalmente el de CLP.

Palabras claves: datos faltantes; series temporales; variables climatológicas

Diseño de experimentos

ANÁLISIS DIALELO MULTIAMBIENTAL APLICADO EN R

JULIA INÉS FERNÁNDEZ¹, DIEGO MARFETÁN MOLINA² y MARÍA LUJÁN FARACE³

¹UNIVERSIDAD NACIONAL DEL NOROESTE DE LA PROVINCIA DE BUENOS AIRES. ²UNIVERSIDAD NACIONAL DE ROSARIO.
³UNIVERSIDAD NACIONAL DEL NOROESTE DE LA PROVINCIA DE BUENOS AIRES – COMISIÓN DE INVESTIGACIÓN CIENTÍFICA DE LA PROVINCIA DE BUENOS AIRES (CIC).

diego.marfetan@gmail.com

Los diseños dialélicos de Griffing (1956) son de gran utilidad para evaluar diferentes aspectos genéticos asociados con las cruzas, que permiten estimar la aptitud combinatoria, importante en el mejoramiento genético cuando se desea comparar el comportamiento de líneas endocríadas con diferentes grados de diversidad genética en combinaciones híbridas. Griffing presentó cuatro métodos en función de la inclusión de los parentales, sus cruzas directas F1 y sus recíprocos, y dos tipos de modelos: de efectos fijos o aleatorios. Singh (1973) desarrolló las sumas de cuadrados para evaluar estos efectos a través de distintos ambientes. Dicho análisis multiambiental está implementado únicamente en programas pagos y en el software AGD desarrollado por el CIMMYT. El objetivo del presente trabajo es construir una función en R que replique los resultados de dichos programas, sea de fácil acceso y pueda ser empleada en programas de mejoramiento genético. Para evaluar su implementación, esta función se aplicó a datos resultantes de un experimento dialélico Método II, Modelo 1 de Griffing entre 5 líneas de maíz (EEA INTA Pergamino) y sus cruzas F1 directas mediante un diseño en bloques anidados en localidades (Pergamino, Junín y Colón). Las variables respuestas consideradas fueron contenido de etanol y proteína, medidas con el método NIRS. Tomando como referencia los resultados proporcionados por AGD, se programó una función en R que permite realizar el análisis dialélico en el contexto de múltiples localidades. Esta función produce como resultados la tabla ANOVA correspondiente, estimaciones de las componentes de habilidades combinatorias general y específica, de la variancia fenotípica y de las heredabilidades en sentido amplio y estricto; además, proporciona estimaciones de los efectos generales y específicos acompañados del ranking de cada parental y cada crusa. Como novedad, se le sumó la opción de indicar el coeficiente de endocría (F) y estimar los parámetros genéticos de acuerdo al valor de F establecido. La implementación en R resultó satisfactoria, logrando reproducir los resultados obtenidos mediante otros paquetes estadísticos. Entre las ventajas de la función propuesta se destacan las posibilidades de aplicar el método a diseños que abarquen varias localidades y de elegir el coeficiente de endocría, opciones hasta ahora no disponibles en R. Si bien actualmente la función puede aplicarse sólo a modelos contemplados por el Método 2 de Griffing, se planea extenderla de forma tal que contemple los restantes diseños.

Palabras claves: dialelo; griffing; multiambiente; r; diseños genéticos; anova

CARACTERIZACION GENOTOXICA DE COMPUESTOS PEPTIDICOS ANTIMICROBIANOS UTILIZANDO EL ENSAYO COMETA EN LEUCOCITOS HUMANOS

GINA DE GIUSTO¹, MARÍA VERÓNICA HÚMPOLA², GEORGINA TONARELLI², MARÍA FERNANDA SIMONIELLO¹, LILIANA ESTER CONTINI³ y OLGA BEATRIZ ÁVILA³

¹CÁTEDRA DE TOXICOLOGÍA, FARMACOLOGÍA Y BIOQUÍMICA LEGAL, FBCB, UNL. ²LABORATORIO DE PÉPTIDOS BIOACTIVOS. DEPARTAMENTO DE QUÍMICA ORGÁNICA, FBCB-UNL. ³DPTO. DE MATEMÁTICA, FBCB - UNL.

lcontini@fbcn.unl.edu.ar

La aparición y rápida propagación de bacterias resistentes a los antibióticos convencionales se ha convertido en una problemática mundial que ha llevado a la búsqueda de nuevos agentes antimicrobianos. En este sentido, los péptidos antimicrobianos (PAs) han surgido como una alternativa para el tratamiento de enfermedades infecciosas. En el diseño de fármacos, la caracterización toxicológica frente a células de mamíferos resulta esencial a fin de evaluar el potencial terapéutico de nuevas drogas. El objetivo de este trabajo fue evaluar las propiedades genotóxicas de compuestos peptídicos antimicrobianos. La genotoxicidad de los 4 péptidos y 9 lipopéptidos se evaluó in vitro, utilizando el ensayo cometa en leucocitos humanos de donantes sanos. Este primer análisis estadístico se realizó con dos experimentos de dos factores con medidas repetidas en uno de ellos. Los factores analizados fueron: concentración con cuatro niveles (medidas repetidas) y el otro, tipo de péptido o lipopéptido. La variable respuesta fue genotoxicidad respecto del valor inicial de cada muestra. Para ambos experimentos se observó que al aumentar la concentración aumentó la genotoxicidad. La interacción concentración*péptido y concentración*lipopéptido, resultó estadísticamente significativa. Del análisis de las comparaciones múltiples considerando todas las combinaciones y de los gráficos de perfil, se concluye que: Lipopéptidos: se observan dos grupos, uno con tres, que es más genotóxico y la interacción se manifiesta en todas las concentraciones; el otro grupo, con seis, menos genotóxico y la interacción se manifiesta solo en las concentraciones bajas. Péptidos: se observa que uno de ellos es más genotóxico en todas las concentraciones; los otros tres, a bajas concentraciones, se comportan de manera similar, a partir del tercer nivel de concentración uno de ellos es tan genotóxico como el primero, mientras que los otros dos, si bien aumenta la genotoxicidad no llegan a los niveles del primero. Esta evaluación permitió seleccionar los compuestos peptídicos que presentan mejor perfil toxicológico para su potencial uso terapéutico.

Palabras claves: péptidos antimicrobi; genotoxicidad; ensayo cometa; lipopéptidos; modelo factorial; medidas repetidas

COMPARACIÓN DE TÉCNICAS NO PARAMÉTRICAS PARA EL ANÁLISIS DE DISEÑOS DE EXPERIMENTOS BIFACTORIALES CON TAMAÑOS MUESTRALES PEQUEÑOS A TRAVÉS DE UN ESTUDIO POR SIMULACIÓN

MARCOS PRUNELLO, MARÍA BELÉN ALLASIA, JUAN JOSÉ IVANCOVICH, SABRINA SILVA QUINTANA, LAURA PISKULIC, HEBE BOTTAI y LILIANA RACCA

ÁREA ESTADÍSTICA Y PROCESAMIENTO DE DATOS, FACULTAD DE CIENCIAS BIOQUÍMICAS Y FARMACÉUTICAS, UNIVERSIDAD NACIONAL DE ROSARIO.

estadistica.fbioyf@gmail.com

El uso de diseños factoriales se ha generalizado en el ámbito científico por su capacidad para detectar efectos de los factores en estudio y sus interacciones. En las ciencias biológicas son frecuentes el pequeño número de observaciones, el incumplimiento de los supuestos requeridos por la técnica clásica y la presencia de observaciones atípicas (outliers). Este trabajo compara cinco técnicas: el método clásico (MC), Aligned Rank Transform (ART), Puri-Sen (PS), Van der Waerden (VDW) y una alternativa robusta (R). Se efectuó un estudio por simulación en contextos donde no se cumple alguno de los supuestos para un diseño bifactorial balanceado con hasta diez unidades experimentales por tratamiento (n). Los datos fueron generados a partir de distintos escenarios: ningún efecto significativo (E1), un efecto principal (E2), ambos efectos principales (E3), sólo efecto interacción (E4), un efecto principal y efecto interacción (E5), todos los efectos significativos (E6). Se evaluó la capacidad de las técnicas para ensayar el efecto de la interacción a través de la estimación de la probabilidad de error de tipo I ($P(EI)$) en los escenarios E1, E2 y E3, y de la potencia en E4, E5 y E6, considerando un nivel de significación $\alpha=0,05$. Cuando se viola el supuesto de normalidad, ART arrojó una $P(EI)$ mayor a 0,05 aproximándose al valor nominal cuando n aumenta. Las restantes técnicas resultaron más conservadoras. Si bien PS y VDW mostraron mayor potencia en E4, ART fue superadora ante la presencia de al menos un efecto principal (E5 y E6) mientras MC y R tuvieron peor desempeño. En el caso de heterocedasticidad, ninguna técnica produjo $P(EI)$ empíricas cercanas a la nominal, con excepción de MC y VDW en E1. Para E4, E5 y E6, ART presentó potencias mayores que el resto. Ante la presencia de outliers para E1 y E2 la $P(EI)$ fue aproximadamente igual a la nominal para todas las técnicas y cualquier n, a excepción de R cuyo desempeño fue inferior. Para E4, E5 y E6, PS, VDW y ART produjeron potencias similares y mayores a las obtenidas con R y MC. En estas dos últimas la potencia no cambió al aumentar n. En conclusión, no fue posible identificar una técnica que mantenga la $P(EI)$ bajo control o provea una potencia satisfactoria en todas las situaciones consideradas. No obstante, en general se observó un mejor desempeño de ART, especialmente frente a la existencia de al menos un efecto principal significativo.

Palabras claves: diseño bifactorial; muestras pequeñas; métodos no paramétricos; simulación; interacción

DESARROLLO SUSTENTABLE DE CULTIVOS HIDROPÓNICOS PARA CONSUMO HUMANO: ANÁLISIS DE TASAS DE CRECIMIENTO

JOSÉ M VARGAS y RICARDO D JUAN

Palabras claves: .

DISTRIBUCIÓN Y PROPORCIÓN DE TESTIGOS EN DISEÑOS CON TESTIGOS A LA PAR

ANDREA PIA¹, MARIANO CORDOBA² y JULIO ALEJANDRO DI RIENZO²

¹FACULTAD DE AGRONOMÍA, UNIVERSIDAD NACIONAL DE LA PAMPA, ARGENTINA. ²FACULTAD DE CIENCIAS AGROPECUARIAS, UNIVERSIDAD NACIONAL DE CÓRDOBA, ARGENTINA.

dirienzo.julio@gmail.com

La primera etapa en el desarrollo de nuevos híbridos y variedades es la evaluación de nuevos genotipos resultantes del cruzamiento de parentales seleccionados. Los programas modernos de selección están enfrentando el síndrome “big data”. Estos programas manejan miles de genotipos en pequeñas parcelas sin repeticiones. La única forma de estimar la varianza residual es incluir réplicas de testigos seleccionados. Debido a que el ensayo puede tener una gran dimensión en el terreno los testigos también tienen el propósito de capturar la variabilidad espacial. El número y ubicación de los testigos es usualmente decidida por razones económicas y operativas. Sin embargo, debido a que el proceso de selección depende de la calidad de las estimaciones del rendimiento de los nuevos genotipos, la decisión de cuántas repeticiones hacer de los testigos y dónde ubicarlos requiere un análisis estadístico. Presentamos los resultados de un estudio de simulación en el que se evalúan varias formas de ubicación sistemática de los testigos, así como la proporción de los mismos en dos tipos de arreglos: uni y bi-dimensionales. Los escenarios simulados fueron generados teniendo en cuenta parámetros típicos que caracterizan estadísticamente la distribución de rendimientos en ensayos típicos de maíz, soja y girasol. Los datos simulados incluyeron errores correlacionados espacialmente. La correlación (modelo exponencial) se especificó mediante la varianza y el rango del proceso gaussiano estacionario subyacente. La simulación se realizó mediante un script escrito en lenguaje R, e implementado alrededor de la librería RandomFields. Para cada escenario y repetición, los datos fueron evaluados suponiendo independencia y también modelando la correlación espacial. La variable evaluada fue la correlación de Spearman entre los efectos genotípicos simulados y los predichos por el modelo ajustado. Los resultados muestran que, como es de esperar, cuanto mayor es la proporción de testigos mejor es la predicción. Sin embargo, como el propósito de los ensayos no es evaluar testigos, y la “ganancia en predicción” es mayor cuando se pasa de una proporción baja a una proporción intermedia, que cuando se pasa de una proporción intermedia a un alta de testigos, la recomendación es utilizar proporciones intermedias (20% del número total de parcelas en arreglos unidimensionales, y 10% en arreglos bidimensionales). En los arreglos bidimensionales, no todos los arreglos son igualmente efectivos.

Palabras claves: mejoramiento vegetal; testigos a la par; correlación espacial

EFICIENCIA RELATIVA DE DISEÑOS EXPERIMENTALES USADOS PARA ESTIMAR REPUESTA A LA FERTILIZACIÓN EN AGRICULTURA DE PRECISIÓN

PABLO PACCIORETTI, MARIANO CÓRDOBA y MÓNICA BALZARINI

UNIVERSIDAD NACIONAL DE CÓRDOBA - CONSEJO NACIONAL DE INVESTIGACIONES CIENTÍFICAS Y TÉCNICAS.

pablopaccioretti@agro.unc.edu.ar

Conocer la respuesta de los cultivos a la fertilización es crucial para optimizar económica y ambientalmente el uso de recursos en agricultura. Esta función puede estimarse a partir de ensayos conducidos con tecnologías de precisión, que hacen posible obtener numerosas repeticiones de la respuesta del cultivo a distintas dosis de fertilización y asociar a cada medición la información sobre su posición en el terreno (georreferencia). Así, se han comenzado a implementar en Argentina, diseños experimentales en campos de productores que incluyen numerosas unidades experimentales (parcelas) aleatorizadas o estratificadas en un número alto de bloques con o sin delimitación de áreas homogéneas dentro del lote. Otro diseño utilizado es el arreglo de parcelas en franjas que atraviesan la variabilidad del lote. La respuesta al interrogante sobre qué diseño de experimento es el más adecuado para estimar funciones de rendimiento, no está del todo clara. El objetivo de este trabajo es evaluar el desempeño de arreglos de parcelas según un diseño completamente aleatorizado (DCA), en bloques completamente aleatorizado (DBCA) y alternativamente en franjas. La comparación se realizó en base a 600 simulaciones de ensayos de fertilización con una estructura de correlación espacial cuyos parámetros fueron determinados a partir de ensayos de fertilización nitrogenada en

cultivo de maíz. En cada realización se adicionó un efecto estadísticamente significativo de zona (área homogénea) con dos niveles. El rendimiento se estimó a partir de una función conocida, cuyos parámetros se fijaron a través de distribuciones de probabilidad estimadas con datos empíricos. Para cada diseño se ajustaron modelos de regresión lineal, con o sin efecto zona y/o contemplando la estructura de correlación espacial del término de error. El DBCA fue el diseño más eficiente tanto en modelos de análisis que incorporaron el efecto de la zona como en aquellos donde se omite este efecto, situación donde la eficiencia relativa del DBCA fue de hasta 5 veces mayor respecto a los otros diseños.

Palabras claves: ensayos; regresión; simulación; estratificación

PROMOVIENDO EL DESARROLLO DEL PENSAMIENTO ESTADÍSTICO EN LA EDUCACIÓN MEDIA A TRAVÉS DEL ABORDAJE DE PROBLEMÁTICAS COMUNITARIAS

*GILDA GARIBOTTI¹, DANIELA ZACHARIÁS¹, CLAUDIA HUAYLLA² y
MAILEN LALLEMENT³*

¹DEPARTAMENTO DE ESTADÍSTICA, CENTRO REGIONAL UNIVERSITARIO BARILOCHE, UN COMAHUE. ²DEPARTAMENTO DE MATEMÁTICA, CENTRO REGIONAL UNIVERSITARIO BARILOCHE, UN COMAHUE. ³GRUPO DE EVALUACIÓN Y MANEJO DE RECURSOS ÍCTICOS, CENTRO REGIONAL UNIVERSITARIO BARILOCHE, UN COMAHUE.

garibottig@comahue-conicet.gob.ar

Los recursos didácticos usualmente utilizados para la enseñanza de estadística están enfocados en cálculos que no permiten comprender el amplio espectro de instancias involucradas en un trabajo estadístico. El desarrollo del pensamiento y razonamiento estadístico es complejo e incluye múltiples aspectos: diseño de experimentos, análisis de datos e interpretación de los resultados en relación a la problemática social o experimental abordada. Este desarrollo se adquiere con mayor profundidad e intuición a través de la realización de proyectos interdisciplinarios donde se aborden problemas que movilicen a los alumnos. El objetivo de este trabajo es proponer una metodología multidisciplinaria, basada en una problemática comunitaria, orientada tanto a contribuir directamente a la formación estadística de alumnos del nivel medio como a brindar herramientas a los profesores. Además, se busca acercar a los alumnos del nivel medio a la universidad, de manera de facilitar su posible inserción en este nivel educativo. Se abordó la temática de la tenencia responsable de perros y su relación con la salud humana. Se diseñó un cuestionario para obtener información demográfica de los perros, características de su cuidado, conocimiento sobre zoonosis y episodios de mordeduras o accidentes de tránsito a causa de perros. En visitas a las escuelas se explicó el procedimiento para realizar un trabajo de investigación, con énfasis en el diseño de experimentos y la obtención de datos. Se capacitó a los alumnos para realizar las entrevistas (cada alumno entrevistó y fue entrevistado por un compañero) y el llenado de la base de datos. Los alumnos de las escuelas concurrieron a la universidad para la presentación del análisis estadístico de la información de la encuesta, las zoonosis más frecuentes en la región y las características principales de la tenencia responsable de perros. A manera de cierre, se realizó un debate sobre alternativas para contribuir a resolver el problema y presentación de propuestas elaboradas por los alumnos de las escuelas. Participaron aproximadamente 200 alumnos de cuarto y quinto año de 4 escuelas de San Carlos de Bariloche, entre mayo y junio de 2017. Se presentó una metodología de acercamiento de la estadística a alumnos de nivel medio recurriendo al estudio multidisciplinario de una problemática social e integrando diferentes materias, facilitando la comprensión de la aplicación de la estadística en contexto y no separada de la realidad. Esta estrategia de trabajo puede ser abordada con temáticas propuestas por los mismos alumnos, incluso referidas a la institución educativa.

Palabras claves: educación en contexto; diseño de experiencias; obtención de datos; estadística descriptiva

Estadística espacial

ANÁLISIS DE COMPONENTES PRINCIPALES GEOGRÁFICAMENTE PONDERADAS UNA PROPUESTA DE ÍNDICE DE BANCARIZACIÓN PARA LA REGIÓN CENTRO (ARGENTINA)

FERNANDO GARCÍA

FACULTAD DE CIENCIAS ECONÓMICAS - UNIVERSIDAD NACIONAL DE CÓRDOBA.

fgarcia.unc@gmail.com

En este trabajo se propone la construcción de un Índice de Bancarización (IB) para la Región Centro (Argentina). Se utiliza información a nivel departamental del Censo 2010 proporcionada por el Instituto Nacional de Estadísticas y Censos e información sobre el sistema bancario provista por el Banco Central de la República Argentina. El indicador considera aspectos referentes a las diferentes dimensiones de análisis: Magnitudes Agregadas, Disponibilidad y Cobertura Geográfica y Utilización y Acceso. Como los datos están georreferenciados, el Análisis de Componentes Principales (PCA) puede verse afectado por los patrones espaciales subyacentes en los datos. En tal sentido, resulta adecuado aplicar PCA Geográficamente Ponderadas (GWPCA) que permite no solo incorporar la dimensión espacial de los datos sino también considerar situaciones donde los datos espaciales no son bien descriptos por un modelo global. Los resultados muestran que las Componentes Principales (CPs) seleccionadas permitirían una mejor visualización de la heterogeneidad espacial y corroborar que GWPCA constituye una estrategia superadora en relación a PCA. De esta manera, se recomienda avanzar en la construcción del IB a partir de las CPs obtenidas a través de esta metodología. Un índice de este tipo es importante en tanto facilita la comprensión de la bancarización y contribuye a que exista un reconocimiento respecto a su trascendencia como elemento que puede apoyar al crecimiento y desarrollo económico. Se destaca la provincia de Santa Fe, exhibiendo un mayor nivel de bancarización. Sigue en importancia la provincia de Córdoba, pero con un comportamiento más heterogéneo que contrasta con el de la provincia de Entre Ríos, que aunque presenta un nivel de bancarización menor exhibe un comportamiento más homogéneo.

Palabras claves: componentes principales; bancarización; índice

APLICACIÓN DE MODELOS LINEALES MIXTOS PARA EVALUAR LA VARIABILIDAD ESPACIAL EN UN SUELO SODICO

NICOLÁS BATTISON¹, MICAELA MANZOTTI¹, JUAN PABLO VELEZ²,
CECILIA MILÁN³ y PAOLA SALVATIERRA¹

¹UNIVERSIDAD NACIONAL DE VILLA MARÍA. ²INTA- MANFREDI. ³UNIVERSIDAD NACIONAL DE VILLA MARÍA.

nicolasjbattiston@gmail.com

La variabilidad de las propiedades edáficas, es entendida como el producto de factores formadores que operan e interactúan en una escala espacial y temporal continua. Describir esta variabilidad a través de técnicas geoestadísticas es de suma importancia para generar estrategias de muestreo, diseños de experimentos, como así también mejorar la capacidad productiva y tomar decisiones sustentables en el uso y manejo de suelos. El objetivo fue caracterizar la distribución espacial de

conductividad eléctrica (CE) y pH en un suelo sódico. El estudio se llevó en un lote de 63,5 ha cercano a la localidad de Arroyo Algodón. Se realizó un muestreo de suelo a través dos grillas de 10x10 m georeferenciadas ubicadas en zonas diferenciadas por altimetría y conductividad eléctrica aparente (CEa) evaluada por un equipo Veris. Por cada punto georeferenciado se determinó CE1-1 (dSm-1) y pH1-2,5 a tres profundidades diferentes 0-20 cm (P1), 20-40 cm (P2) y 40-60cm (P3). Se ajustaron diferentes modelos lineales mixtos con estructuras de correlación espacial y se utilizó el criterio de Akaike para seleccionar el mejor ajuste. Posteriormente con los parámetros estimados se calcularon índices de varianza estructural relativa (RSV) y mapas de predicción. Las funciones de semivarianza espacial para ambas grillas fueron principalmente Lineal, Exponencial y Gaussiana. Los rangos de distribución tanto para pH1-2,5 como para CE1-1 oscilaron entre 10 y 18 m. Estos valores no se vieron modificados en gran medida ni por la profundidad de la muestra, ni por la altimetría y CEa. En cuanto al RSV, fueron considerados altos, medios y bajos variando según la propiedad y profundidad. Se pudo observar que, cuando el grado de estructuración espacial fue alto para CE 1:1 para pH 1:2,5 fue medio y bajo, en las mismas profundidades analizadas. A través del análisis de mapas de predicción se podría concluir que ambas propiedades presentan una estructura de variación tipo vertical y horizontal.

Palabras claves: distribución espacial; suelos alcalinos; geoestadística moder

COMPARACIÓN DE ALGORITMOS DE CLASIFICACIÓN NO SUPERVISADA PARA EL ESTUDIO DE COBERTURAS FORRAJERAS EXTENSIVAS

DIEGO CARCEDO y MIGUEL NOLASCO

CÁTEDRA DE ESTADÍSTICA Y BÍOMETRÍA - FACULTAD DE CIENCIAS AGROPECUARIAS - UNIVERSIDAD NACIONAL DE CÓRDOBA.

dncarcedo@gmail.com

El conocimiento de los tipos de cobertura de vegetación en tierras destinadas a la producción animal extensiva permite una mejor administración de la biomasa forrajera disponible. Tradicionalmente la caracterización de la oferta forrajera se realiza en base a muestreos a campo. En los últimos años, el acceso a datos provenientes de sensores remotos y a servicios para su procesamiento en línea, permiten obtener series temporales de índices de vegetación para múltiples sitios en el dominio de estudio. El objetivo del presente trabajo fue comparar el comportamiento de los métodos de conglomerados K-means (no jerárquico) y UPGMA (jerárquico) en el contexto de datos geoespaciales temporales. El área de estudio (noroeste de la Provincia de Córdoba), abarcó una superficie rectangular de 407 km². En primera instancia, utilizando el algoritmo de segmentación denominado Meanshift, se confeccionaron objetos espaciales que respetaran la homogeneidad de las distintas coberturas presentes en el terreno. Empleando la plataforma web Google Earth Engine (GEE) se construyó una serie temporal del producto Índice de Vegetación Diferencial Normalizado (IDVN) (2013-2015 con 2 o más índices por mes). Las colecciones de imágenes satelitales Landsat 7, Landsat 8 y Sentinel 2, fueron usadas para cada objeto y tipo de cobertura. Los valores de la serie temporal resultante fueron resumidos estadísticamente (media y DE) por año y estación. Posteriormente se calculó una matriz de distancias entre los objetos, utilizando la métrica Euclídea, sobre estas propiedades de los datos. Esta matriz de distancia fue el input de los algoritmos K-means y UPGMA. Para identificar el número apropiado de grupos existentes en la estructura de los datos, se usaron los índices Ancho de silueta (IAS), Dunn (ID) y Conectividad (IC). Los valores obtenidos evidenciaron que ambos algoritmos identificaron una partición apropiada en dos grupos (K-Means: 0,76; 0,98; 1 y UPGMA: 0,76; 0,98; 1 para IAS, ID y IC respectivamente). Este agrupamiento puede explicarse por la diferencia en el patrón temporal de la serie de datos. Una de las clases presentó valores estables y periódicos en el tiempo, mientras que la restante exhibió una mayor inestabilidad. El uso de propiedades estadísticas de ventanas del tamaño de una estación en series de NDVI permitió contemplar

la estructura temporal subyacente en los datos.

Palabras claves: k-means; upgma; imágenes satelitales; google earth engine

DESARROLLO DE UN NUEVO MAPA DE RIESGO EPIDEMIOLÓGICO PARA LEPTOSPIROSIS EN LA CIUDAD DE ROSARIO

LUCRECIA CORALLO¹, LAURA BALPARD¹, XIMENA PORCASI², MARIO LAMFRI², ANDREA FRETES CHÁVEZ¹ y DIEGO LÓPEZ³

¹SISTEMA MUNICIPAL DE EPIDEMIOLOGÍA. SECRETARÍA DE SALUD PÚBLICA. MUNICIPALIDAD DE ROSARIO.. ²INSTITUTO DE ALTOS ESTUDIOS "MARIO GULICH". COMISIÓN NACIONAL DE ACTIVIDADES ESPACIALES.. ³ÁREA DE SENsoRES REMOTOS. ESCUELA DE AGRIMENSURA. FACULTAD DE CIENCIAS EXACTAS, INGENIERÍA Y AGRIMENSURA. UNIVERSIDAD NACIONAL DE ROSARIO..

lbalparda@hotmail.com

Este trabajo describe mejoras de un mapa preliminar de riesgo para leptospirosis de la ciudad de Rosario, considerando un brote ocurrido en el año 2007 luego de un evento de inundación en la ciudad. El mapa preliminar fue presentado en publicaciones previas (Corallo et.al., 2010). Para la generación del mapa que aquí se presenta, se consideraron 21 espacios territoriales donde se localizaron casos confirmados de leptospirosis ocurridos entre el 04/04/2007 al 16/05/2007 (grupo A) y 70 nuevos espacios territoriales elegidos por muestreo simple al azar donde no se localizaron casos (grupo B). Se incluyó la correspondiente corrección atmosférica y calibración radiométrica a las imágenes satelitales Landsat 5TM (17/03/2007, 18/04/2007 y 20/05/2007). Se calcularon los índices de vegetación, agua y suelo, y sus correspondientes diferencias. Se crearon áreas de influencia hasta 11.000 metros de las zonas de interés (inundadas/anegadas, vegetación y ferrocarriles). Específicamente, los aspectos mejorados o incorporados para el nuevo mapa fueron: a) modificación del criterio para la construcción de las áreas de influencia de las zonas de interés; b) incorporación de la corrección atmosférica y calibración radiométrica, en cada una de las imágenes (Chander et. al., 2009); c) profundización del análisis descriptivo, adición del análisis de componentes principales y ajustes en el análisis de regresión logística. Resultados: Análisis descriptivo con estadísticas básicas y gráficos boxplot, según grupos A y B. En el análisis de componentes principales se identificó, para cada una de las imágenes y sus índices, que las dos primeras componentes explicaron más del 80% de la variabilidad total. Según los ajustes en el análisis de regresión logística, las variables de mayor poder predictivo fueron: la distancia a las zonas inundadas/anegadas, la primer componente principal de la imagen del mes de mayo y la segunda de la imagen de marzo (sensibilidad 52%; especificidad 90%). Considerando el modelo de regresión, las variables y sus coeficientes, se construyó el nuevo mapa de riesgo para leptospirosis de la ciudad de Rosario para el año 2007, con un procedimiento optimizado más riguroso que permitió salvar las dificultades preliminares. El trabajo muestra los mapas preliminar y actualizado en el que se evidencian mejoras en la especificidad de predicción del riesgo. El procedimiento de generación del nuevo mapa, puede replicarse a futuro ante la ocurrencia de intensas lluvias que pudieran provocar la inundación o el anegamiento de terrenos, importante factor de riesgo para la aparición de un nuevo brote de leptospirosis en la ciudad.

Palabras claves: mapa; riesgo; epidemiológico; leptospirosis; imágenes satelitales

EVALUACIÓN DE LA DELIMITACIÓN DE ZONAS HOMOGÉNEAS EN SUELOS SÓDICOS INCLUYENDO INFORMACIÓN ESPACIAL

*MARTIN PRARIZZI¹, NICOLAS BATTISTON¹, MICAELA MANZOTTI¹,
CECILIA MILÁN² y PAOLA SALVATIERRA¹*

¹UNIVERSIDAD NACIONAL DE VILLA MARÍA. ²UNIVERSIDAD NACIONAL DE VILLA MARÍA.

martinprari@gmail.com

En el marco de este trabajo se realizó un ensayo con la finalidad de comparar los ANÁLISIS DE COMPONENTES PRINCIPALES CLÁSICO (PCA) y ANÁLISIS DE COMPONENTES PRINCIPALES ESPACIALES (MULTISPATI-PCA). Se recolectaron datos en un lote de 41 ha. con valores geo-referenciados de conductividad eléctrica aparente CEa (mS m-1) en dos profundidades 0-30 cm (CE30) y 0-90cm (CE90) y altimetría (m) a través de un sensor de arrastre (modelo 3100, VerisTech Inc., EEUU) y un DGPS (Raven P300 con corrección de señal Omnistar, de precisión submétrica) que además les asocio coordenadas geográficas (sistema Lat Lon-WGS84). Como punto de partida se llevó a cabo la transformación de coordenadas y depuración de datos outliers. Para la delimitación de las zonas homogéneas en primer lugar se realizaron MULTISPATI-PCA Y PCA para asociar los datos de las diferentes variables relevadas. Los resultados obtenidos por ambos métodos se compararon a través de la pérdida de inercia (varianza espacial vs varianza total) y el aumento de la información espacial (Índice de Moran de MULTISPATI-PCA vs Índice de Moran del PCA). Posteriormente con las primeras dos componentes espaciales y clásicas se realizaron clasificaciones usando el algoritmo de FUZZY K-MEANS y con el agrupamiento obtenido, en cada caso, se realizó la visualización gráfica utilizando un suavizado. Para la determinación del número óptimo de grupos se utilizaron los índices de “Xie.Beni”, Fukuyama Sugeno, Coeficiente de partición y entropía de partición. Los resultados mostraron que con MULTISPATI-PCA se explicó una menor proporción de varianza acumulada en la primer componente respecto de PCA, mientras que en la segunda componente esta relación se invierte. En cuanto al índice de Moran calculado sobre cada una de las componentes, se observó que la autocorrelación no fue más alta al incluir información espacial. En cuanto a la clasificación, ambos casos, los resultados obtenidos fueron dos zonas bien delimitadas y homogéneas dentro del lote. Esto indica que, en esta aplicación, la introducción de información espacial no proporciona una ventaja adicional.

Palabras claves: multispati-pca; fuzzi-kmeans; pca

USO DE TÉCNICAS GEOESTADÍSTICAS PARA CUANTIFICAR EL EFECTO DE LA INTENSIDAD DE TRÁNSITO SOBRE LA DISTRIBUCIÓN HORIZONTAL Y VERTICAL DE LA COMPACTACIÓN EN SUELOS AGRÍCOLAS

C.A. ALESSO¹, P.A. CIPRIOTTI², M.J. MASOLA³, ME CARRIZO³, DL ANTILLE⁴ y SC IMHOFF³

¹DPTO. DE CIENCIAS BÁSICAS, FACULTAD DE CIENCIAS AGRARIAS, UNIVERSIDAD NACIONAL DEL LITORAL. (FCA-UNL). ²DPTO. DE MÉTODOS CUANTITATIVOS Y SISTEMAS DE INFORMACIÓN, FACULTAD DE AGRONOMÍA, UNIVERSIDAD DE BUENOS AIRES (FAUBA). INSTITUTO DE INVESTIGACIONES FISIOLÓGICAS Y ECOLÓGICAS VINCULADAS A LA AGRICULTURA. ³DPTO. DE BIOLOGÍA VEGETAL, FACULTAD DE CIENCIAS AGRARIAS, UNIVERSIDAD NACIONAL DEL LITORAL. (FCA-UNL). ⁴UNIVERSITY OF SOUTHERN QUEENSLAND, NATIONAL CENTRE FOR ENGINEERING IN AGRICULTURE, TOOWOOOMBA, QLD, AUSTRALIA..

calesto@fca.unl.edu.ar

El tránsito controlado es un sistema de mecanización en donde la maquinaria agrícola transita dentro del lote únicamente por sendas predeterminadas para reducir la superficie compactada por las ruedas y maximizar las condiciones del suelo para la exploración radical y transitabilidad. La magnitud y distribución espacial de la compactación, estimada por la resistencia mecánica a la penetración (RP), en torno a las sendas podría explicar la respuesta de los cultivos en las hileras próximas a las huellas. El objetivo de este trabajo fue modelar la distribución espacial de la RP en el plano vertical en torno a las sendas de circulación sobre un suelo Argiudol Típico Serie Rafaela de la localidad de Aurelia, Santa Fe (Argentina). Se simularon cuatro intensidades de tránsito (0, 6, 12 y 18 pasadas de un tractor Pauny 230C) en parcelas de 6 m x 30 m, siguiendo un diseño en bloques completos al azar con tres réplicas. Previo a la siembra de soja y trigo, en cada parcela se registró la RP de los primeros 30 cm del suelo cada 2 cm de profundidad (dirección vertical) insertando un penetrómetro digital de punta cónica cada 5,5 cm en transectas de 110 y 187 cm perpendiculares a las sendas (dirección horizontal). La distribución espacial de los datos en la grilla vertical XZ de 5,5 cm x 2 cm se modeló mediante técnicas geoestadísticas y se obtuvieron superficies interpoladas por kriging universal. Se estimó la proporción de área con valores de RP ≥ 2 MPa (límite para el crecimiento radical) en las zonas bajo huella y entre huellas la cual se utilizó como variable respuesta ajustando un modelo con medidas repetidas en el espacio. En todos los casos, los valores de RP mostraron tendencia en ambas direcciones la cual explicó 50-84% de la variación total. La variación residual mostró una estructura espacial esférica con alcances de 15 a 36 cm. Previo a la siembra de la soja no se detectaron diferencias significativas en la proporción de área compactada. En cambio, previo a la siembra del trigo se detectaron diferencias significativas entre las intensidades y entre zonas. La proporción de área compactada siguió una tendencia cuadrática en función del número de pasadas ($Y = 0.25 + 0.035X - 0.00176X^2$). El 45% del área bajo la huella tuvo RP ≥ 2 MPa mientras que sólo el 22% mostró compactación fuera de la huella.

Palabras claves: compactación; geoestadística; resistencia mecánica; tránsito controlado; variabilidad espacial

VARIABILIDAD ESPACIO-TEMPORAL DE LAS PRECIPITACIONES AL OESTE DE LA REGIÓN PAMPENA – IMPLICANCIAS PARA LA PRODUCCIÓN DE GRANOS

TERESA BOCA¹ y PABLO CIPRIOTTI²

¹INSTITUTO NACIONAL DE TECNOLOGÍA AGROPECUARIA (INTA). DPTO. DE MÉTODOS CUANTITATIVOS Y SISTEMAS DE INFORMACIÓN, FACULTAD DE AGRONOMÍA, UNIVERSIDAD DE BUENOS AIRES (FA-UBA). ²DPTO. DE MÉTODOS CUANTITATIVOS Y SISTEMAS DE INFORMACIÓN, FACULTAD DE AGRONOMÍA, UNIVERSIDAD DE BUENOS AIRES (FA-UBA) INSTITUTO DE INVESTIGACIONES FISIOLÓGICAS Y ECOLÓGICAS VINCULADAS A LA AGRICULTURA.

tereboca@gmail.com

Durante las últimas décadas ha habido un corrimiento de la frontera agrícola hacia al oeste de la Región Pampeana sobre áreas más limitantes para los cultivos en secano. Este desplazamiento de la agricultura se dio hacia regiones de mayor variabilidad inter-anual en las precipitaciones, como es el este de la provincia de La Pampa, con un claro impacto sobre el rinde y la seguridad de cosecha de los principales cultivos. El objetivo de este trabajo fue describir la variabilidad espacio-temporal de las precipitaciones en dicha zona, que ocupa más de 130,000 km². Se trabajó con series históricas de precipitaciones mensuales de 85 sitios con registros entre 50-94 años en el período 1921-2015 de la red provincial pampeana del agua. Se definieron 16 variables de respuesta por sitio: precipitación anual, precipitaciones acumuladas para tres cultivos: trigo (jun-dic), maíz (sep-mar) y soja (nov-abr), y 12 precipitaciones mensuales. Para cada variable respuesta se describió la variabilidad espacio-temporal a través de: 1- estimar las tasas de cambio a través del tiempo mediante análisis de regresión lineal en cada sitio; 2- modelar la estructura de correlación espacial del estimador de la pendiente temporal a partir de variogramas; Y 3- mapear la estructura espacial de la pendiente

en la región a través de kriging ordinario. De las 1360 regresiones analizadas (85 sitios x 16 precipitaciones) el 28% presentaron pendientes significativas ($\alpha=5\%$), sin embargo los resultados dependieron mucho de la variable de precipitación considerada. Tanto la precipitación anual como la de los cultivos de verano (maíz y soja) tuvieron más del 75% de las pendientes significativas, aunque con valores de R^2 bajos ($<30\%$) y estimadores puntuales entre 1,2–4,6 mm/año. Esto se debió a que los cambios más consistentes en las precipitaciones mensuales a través del tiempo ocurrieron en los meses de verano y otoño (Ene–Abr), sosteniendo el corrimiento de la frontera agrícola. El análisis geo-estadístico mostró una correlación espacial estructurada variable en las tasas de cambio temporales (20–80%), siendo mayor para las precipitaciones mensuales de Enero, Febrero, Marzo y el periodo del cultivo de soja, con alcances entre 50–150 km. Los resultados encontrados están en línea con la tercera comunicación nacional de la República Argentina para la Convención de las Naciones Unidas sobre el Cambio Climático (IPCC) que afirma que los mayores cambios se registraron en el este del país y con las variaciones más importantes en algunas zonas semiáridas.

Palabras claves: autocorrelación; climatología; geoestadística; precipitaciones; variograma; producción de granos

Inferencia estadística

ANÁLISIS ESTADÍSTICO PARA EVALUAR UN MÉTODO DE ACREDITACIÓN DE IDENTIDAD EN RECIÉN NACIDOS

MARIA DE LA PAZ GUILLOUN¹, LILIANA AMALIA GARCIA¹ y MARIA DEL CARMEN COVAS²

¹UNIVERSIDAD NACIONAL DEL SUR. ²HOSPITAL PRIVADO DEL SUR, BAHÍA BLANCA.

mguillon@criba.edu.ar

La correcta identidad de las personas desde su nacimiento es un derecho inalienable que fija nuestra Constitución: “Todo niño nacido vivo o muerto y su madre deben ser identificados de acuerdo con las disposiciones de la ley 24540”. Según las instituciones que asisten nacimientos, para identificar al recién nacido se siguen diferentes procedimientos, los cuales se aplican simultáneamente en la madre y su hijo antes de abandonar la sala de partos (incluye quirófano). La colocación y mantenimiento de pulseras de identificación es la conducta más utilizada. Sin embargo, es un método que puede resultar no confiable en algunos casos, dado que en la práctica se sabe que la pulsera puede removese o desprenderse de manera natural del recién nacido. Dada la importancia que reviste entonces la pulsera en la identificación de un recién nacido, surgió en enfermeras y médicos del Servicio de Neonatología del Hospital Privado del Sur de Bahía Blanca, Argentina, la inquietud por averiguar cuán confiable es la pulsera como elemento de identificación. Con este fin, el Centro de Estudios de Calidad Total fue convocado por dicho Servicio para realizar el asesoramiento estadístico que el problema requería. Conjuntamente se decidió realizar un estudio de observación del comportamiento de la pulsera, prospectivo, tipo cohorte, aleatorizado, en 914 recién nacidos en el citado hospital, lo que se completó a lo largo de diez meses entre 2013 y 2014. En este trabajo se describe la planificación de las actividades requeridas para el citado estudio; la organización del equipo que lo llevó a cabo; modalidad, capacitación brindada e inconvenientes surgidos en la recolección de los datos; la base de datos diseñada; las pruebas realizadas y conclusiones alcanzadas. Se observaron, entre otras variables, el lugar de colocación de la pulsera (pierna o brazo) en el recién

nacido, la presencia de la pulsera al momento del alta, las causas y la incidencia del familiar o personal de enfermería en caso de su ausencia. El análisis estadístico fue realizado mediante el software SPSS 15.0 y Statgraphics Centurion XV. La conclusión más relevante fue que la colocación de la pulsera de identificación no resulta un método seguro para la identificación del recién nacido. La tercera parte de los mismos, al momento del alta, no la tenía en el lugar (brazo o pierna) donde había sido colocada. La salida espontánea fue la causa más frecuente de su ausencia, con una mayor permanencia en la pierna.

Palabras claves: inferencia estadísti; identificación del r; pulsera de identific

CARACTERIZACIÓN DE LOS ESTUDIANTES DE BIOESTADÍSTICA Y RELACIÓN CON LA CONDICIÓN FINAL. CARRERA DE NUTRICIÓN. FACULTAD DE SALUD. UNSA

SERGIO LEONARDO FONTEÑEZ, SILVIA ROCIO ECHALAR y MARIA DEL CARMEN HERRERA

CÁTEDRA DE BIOESTADÍSTICA. FACULTAD DE CIENCIAS DE LA SALUD. UNIVERSIDAD NACIONAL DE SALTA.

slfsergioslf@gmail.com

INTRODUCCIÓN: Toda Cátedra tiene como objetivo el conocer a los sujetos del aprendizaje y favorecer el éxito académico que se verá reflejado en la condición final del cursado. **OBJETIVOS:** Caracterizar a los estudiantes que cursaron la asignatura de Bioestadística. Relacionar la condición de cursado de Estadística Descriptiva con la condición final de Bioestadística. **DESARROLLO:** Se realizó un estudio de corte transversal, descriptivo y analítico. Se aplicó una encuesta online en la plataforma Moodle, la muestra estuvo constituida por 99 estudiantes, se evaluaron variables sociales, de percepción y rendimiento académico. Se aplicaron análisis descriptivos y pruebas inferenciales con InfoStat v.2012. **RESULTADOS:** El promedio de edad fue 21.34 ± 3.13 años, con una edad mínima de 18 y máxima de 41 años. El 89% de los estudiantes fueron mujeres. Del estado Civil el 95% fue soltero. Con respecto a la procedencia se observó que el 64.29% procede de Salta Capital y un 16,33% de la Provincia de Jujuy. En el ciclo lectivo 2017 se registró que el 55.56% eran Ingresantes, de los cuales el 87% finalizó Estadística Descriptiva, y entre los recursantes el 84%. Los temas que resultaron más complejos en su aprendizaje fueron Correlación y Regresión (41.41%) y Pruebas de Hipótesis (27.27%), cabe destacar que un 15% de los estudiantes manifestaron ninguna dificultad. En cuanto al tema que consideran de mayor utilidad en su profesión encontramos que el 25.25% considera a todos de igual importancia. El 69.70% califica como normal su nivel de conocimiento en Estadística Inferencial una vez finalizado el programa. Al finalizar el cursado 49.49% promocionó, el 38.38% regularizó y el 10.10% no regularizó. La condición final de los estudiantes de Bioestadística (Promocionó y Regularizó) es independiente de tener finalizada Estadística Descriptiva ($p(v)= 0.1533$). **CONCLUSIONES:** Las características de los estudiantes de Bioestadística obtenidas vienen con una tendencia similar en los últimos años, tener o no finalizada estadística descriptiva no condiciona un buen rendimiento del estudiante.

Palabras claves: bioestadística; estadística descript; estudiantes de nutri; condición final

DIFERENCIAS EN NIVELES DE COLESTEROL HDL ENTRE NIÑOS KOYAS Y DE ESTADOS UNIDOS

*CLAUDIA MOLINARI¹, CLAUDIO GONZALEZ², GUSTAVO MACCALLINI³,
MARIANA HIDALGO³ y VALERIA HIRSCHLER²*

¹FACULTAD DE FARMACIA Y BIOQUÍMICA-UBA. ²UBA. ³LABORATORIOS HIDALGO.

cmolinari@ffyb.uba.ar

La dislipidemia es uno de los factores de riesgo cardiovascular en niños, sin embargo muy pocos estudios se llevan a cabo sobre poblaciones indígenas viviendo en condiciones muy diferentes a las urbanas. El presente estudio se basó en la observación de niños de la población de San Antonio de los Cobres (SAC), localizado a una altura promedio de 3750 m sobre el nivel del mar, con una temperatura media anual de menos de 10°C. Los habitantes de esta región muestran diferencias con los individuos de otras localidades situadas a menor altura debido a la hipoxia crónica asociada con la altitud. Un estudio llevado a cabo en USA sugiere que la reducción de la temperatura en alturas elevadas puede llevar a la pérdida de peso. Consistentemente con ello, un estudio previo sobre la misma población de SAC mostró que tienen una reducción significativa en los niveles de hemoglobina y en la prevalencia de obesidad, respecto de niños viviendo al nivel del mar. En este caso se aplicaron métodos estadísticos apropiados para comparar los niveles de HDL en estos niños y los niños americanos incluidos en el estudio del National Center for Health Statistics (NHANES), que constituye un estudio de referencia internacional para parámetros poblacionales. La muestra de NHANES es una muestral nacional representativa de la población norteamericana. La muestra en que se basa el presente estudio consistió de 1232 niños de la localidad de San Antonio de los Cobres, con edad entre 4 y 14 años, que fueron comparados estadísticamente con los 1451 niños americanos incluidos en NHANES 2011-2012. Se testeó la diferencia de medidas antropométricas, la prevalencia de sobrepeso y obesidad, y la prevalencia de HDL por debajo del valor admitido como normal. Se comparó también el nivel de HDL según categorías de MIC (masa de índice corporal). Se aplicó regresión logística a fin de evaluar la diferencia de riesgo de bajo HDL según origen del niño, controlando por variables antropométricas. A partir de los análisis realizados se concluyó que los niños de SAC resultan significativamente más delgados y bajos que los niños americanos. La prevalencia de HDL bajo es significativamente más alta que en sus pares de USA. Los niños de SAC tienen 9,5 veces más de chance de tener HDL por debajo del punto de corte considerado como normal, que los niños americanos, ajustando por posibles confusores.

Palabras claves: hdl; dislipidemia; nhanes; sac; prevalencia; regresión logística

EFICACIA Y SEGURIDAD DE UNA FORMULACIÓN INNOVADORA DE LATANOPROST 0,005% LIBRE DE CLORURO DE BENZALCONIO PARA EL TRATAMIENTO DEL GLAUCOMA DE ÁNGULO ABIERTO

*CELINA LOGIOCO¹, VIRGINIA ZANUTIGH¹, MARÍA SILVIA PASSERINI¹,
MARÍA DE LOURDES RODRIGUEZ¹, MARÍA JOSÉ CASTRO² y MYRIAM
NUÑEZ²*

¹CENTRO DE OJOS QUILMES. ²UNIVERSIDAD DE BUENOS AIRES. FACULTAD DE FARMACIA Y BIOQUÍMICA. DEPARTAMENTO DE FISICOMATEMÁTICA. CÁTEDRA DE MATEMÁTICA.

myriam@ffyb.uba.ar

El Glaucoma de ángulo abierto, también conocido como el ladrón de la vista es, junto al glau-

coma de ángulo cerrado, uno de los dos tipos de glaucoma existentes. Es una enfermedad del ojo cuyo principal factor de riesgo es la elevación de la presión intraocular. El cloruro de benzalconio (BAK) se utiliza en formulaciones de análogos prostaglandínicos debido a su acción adyuvante y conservante. Sin embargo, su acción conservante tiene efectos tóxicos conocidos sobre la superficie ocular, causando sequedad ocular y malestar a largo plazo. El objetivo de este estudio es evaluar la eficacia y tolerabilidad de una nueva emulsión latanoprost 0.005% (LEBAK-free) respecto al a la solución tradicional de Latanoprost 0.005% con BAK (LSBAK). Se llevó a cabo un estudio prospectivo durante 12 semanas. Un total de 16 pacientes (32 ojos) con glaucoma de ángulo abierto primario que fueron tratados con LSBAK por más de 6 meses. Luego se procedió a la suspensión de la aplicación de LSBAK para comenzar con el uso de LEBAK-free. La eficacia y la tolerabilidad de las formulaciones fueron evaluadas: 1 momento de suspender el tratamiento con LSBAK y luego de 12 semanas de recibido el nuevo producto. El valor de la presión intraocular (PIO) se midió ($n = 32$ ojos) al momento de comenzar el tratamiento con LEBAK-free (semana 0), luego a las 4, 8 y 12 semanas. Se investigó la ocurrencia de efectos adversos y tolerabilidad al nuevo producto. Se analizó la hipótesis de independencia entre ambos ojos. Para la variable PIO se estudió la existencia de diferencias significativas a lo largo de las semanas, Análisis de la Varianza con Medidas Repetidas. Para esto se analizó el supuesto de normalidad de la variable en estudio que se requiere para la aplicación del método. Se comprobó que la variable seguía una distribución normal (Test de Shapiro Wilks, $p = 0,359$). El ANOVA reflejó que, existen diferencias significativas entre las medias de los 4 grupos ($p = 0,005$). Las comparaciones a posteriori mostraron la existencia de diferencias significativas entre las medias de la PIO para las semanas 1 y 4 (Método de Bonferroni). Se concluyó que la nueva fórmula de Latanoprost 0.005% en emulsión sin BAK garantiza una distribución homogénea en la superficie ocular, asegurando como mínimo la misma eficiencia en reducir la PIO que la solución tradicional, mejorando la tolerabilidad local y con pocos efectos adversos en la superficie ocular.

Palabras claves: glaucoma de ángulo a; superficie ocular; eficacia y tolerabil; presión intraocular

ESTUDIO COMPARATIVO DEL TEGUMENTO DE TRES VARIEDADES DE ARVEJAS COMERCIALIZADAS EN LA CIUDAD DE HUMAHUACA. ANÁLISIS ESTADÍSTICO PARA SU DIFERENCIACIÓN MICROSCÓPICA

JUDITH MONTENEGRO BRUSOTTI¹, GRACIELA BASSOLS² y MYRIAM NUÑEZ³

¹CÁTEDRA DE FARMACOBOTÁNICA, FACULTAD DE FARMACIA Y BIOQUÍMICA, UNIVERSIDAD DE BUENOS AIRES. CÁTEDRA DE MATEMÁTICA, FACULTAD DE FARMACIA Y BIOQUÍMICA, UNIVERSIDAD DE BUENOS AIRES. ²CÁTEDRA DE FARMACOBOTÁNICA, FACULTAD DE FARMACIA Y BIOQUÍMICA, UNIVERSIDAD DE BUENOS AIRES. ³CÁTEDRA DE MATEMÁTICA, FACULTAD DE FARMACIA Y BIOQUÍMICA, UNIVERSIDAD DE BUENOS AIRES.

jmontenegro@ffyb.uba.ar

En la ciudad de Humahuaca, localizada en la Provincia argentina de Jujuy, se encuentran comercializándose a granel tres variedades de arvejas. Macroscópicamente son bien diferentes, pero a nivel microscópico la diferencia es válida si se realizan métodos micrométricos y estadísticos para comparar largo y ancho de algunas esclereidas que componen el tegumento de las semillas. La necesidad de una identificación a nivel microscópico surge de que su uso y/o comercialización suele estar en forma trozada o de harinas. El objetivo del presente trabajo es poder identificar, a través de la diferenciación microscópica del tegumento la variedad de la cual proviene la arveja. La metodología empleada para la preparación del material en estudio consiste en hidratar las arvejas, retirar los tegumentos y realizar con ellos un disociado fuerte para poder separar las células. Se observa bajo microscopio óptico y con un ocular graduado se mide el tamaño de las esclereidas. El resultado del análisis mostró 2 tipos de esclereidas (esclereidas columnares y osteoesclereidas) que se encontraban en las tres variedades de arvejas, algunas con un amplio rango de tamaño que

hace necesario un tratamiento estadístico para validar diferencias. Al realizar el análisis estadístico (Análisis de la varianza) no se evidencia diferencia significativa entre las tres variedades de arvejas, columnar pero si se obtienen diferencias significativas entre las tres variedades del otro tipo celular, las osteoesclereidas. Se concluye a partir de este resultado que ante una presentación comercial de harina o pequeños trozos de arveja, es posible determinar la variedad de la que proviene.

Palabras claves: arveja; tegumento; esclereida; análisis estadístico

EXPLORACIÓN DE EFECTOS GENÉTICOS EN GENOTIPOS DE PRIMER Y SEGUNDO CICLO DE TOMATE

*PAOLO CACCHIARELLI¹, VLADIMIR CAMBIASO¹, JAVIER PEREIRA DA
COSTA¹, ELIZABETH TAPIA², GUSTAVO R. RODRÍGUEZ¹ y GUILLERMO R.
PRATTA¹*

¹IICAR (INSTITUTO DE INVESTIGACIONES EN CIENCIAS AGRARIAS DE ROSARIO), CONICET/UNR. ²CIFASIS (CENTRO INTERNACIONAL FRANCO ARGENTINO DE CIENCIAS DE LA INFORMACIÓN Y SISTEMAS), CONICET/UNR.

cacchiarelli@iicar-conicet.gob.ar

Los programas de mejoramiento genético de tomate (*Solanum lycopersicum* L.) pueden disponer de Genotipos de Primer Ciclo (GPC, progenitores cultivados y exóticos de constitución genotípica homocigota y sus híbridos, F1 heterocigotas) y Genotipos de Segundo Ciclo (GSC): nuevas líneas homocigotas derivadas de la generación segregante de los GPC mediante un proceso de Selección Artificial (SA), que presentan por Recombinación Genética (RG) características deseables de ambos progenitores, y sus respectivas F1. Por otro lado los híbridos suelen presentar Heterosis (H) o diferencias con respecto a la media entre sus progenitores para caracteres cuantitativos. También pueden encontrarse en ellos efectos recíprocos según la Dirección de Cruzamiento (DC), es decir cuál progenitor se usó como madre y cuál como padre. El objetivo fue explorar, mediante la comparación de GPC y GSC de tomate, los efectos de SA, RG, H y DC sobre caracteres de calidad de fruto. Se evaluaron los GPC: cultivar Caimanta de *S. lycopersicum* (C), LA0722 de *S. pimpinellifolium* (P) y sus F1 recíprocas (CxP y PxP), y los GSC: ToUNR1 (L1) y ToUNR18 (L18) y sus F1 recíprocas (L1xL18 y L18xL1). En invernadero se cultivaron 8 plantas de cada genotipo (n=64), en cuyos frutos se evaluaron peso, vida poscosecha, forma, sólidos solubles (SS), pH y firmeza. Para cada carácter cuantitativo se compararon todos los genotipos mediante ANOVA a un criterio de clasificación. Para explorar los diferentes efectos, se plantearon los siguientes contrastes ortogonales: 1- SA y RG: C, P, CxP, PxP vs L1, L18, L1xL18, L18xL1; 2- H en GPC: CxP, PxP vs C, P; 3- H en GSC: L1xL18, L18xL1 vs L1, L18; 4- DC en GPC: CxP vs PxP; 5- DC en GSC: L1xL18 vs L18xL1. Para todos los caracteres, la distribución fenotípica ajustó a la normal (excepto para peso, que fue transformado mediante logaritmo natural) y se encontraron diferencias significativas entre todos los genotipos. Respecto a los diferentes efectos explorados, los de SA y RG fueron significativos para todos los caracteres excepto para SS. H en GPC fue significativa para todos los caracteres excepto para pH; mientras que H en GSC solo fue significativa para peso. DC fue no significativa para todos los caracteres en GPC y significativa sólo para vida poscosecha y peso en GSC. Se concluye que comparando GPC y GSC mediante contrastes ortogonales, fue posible determinar los efectos de SA, RG, H y DC sobre caracteres cuantitativos de calidad de fruto.

Palabras claves: selección artificial; recombinación genética; heterosis; dirección de cruzamiento

INTERVALOS DE CONFIANZA PARA LA RAZÓN COSTO-EFECTIVIDAD: UNA COMPARACIÓN DE TÉCNICAS PARAMÉTRICAS Y NO PARAMÉTRICAS MEDIANTE SIMULACIÓN

PABLO COTTET y NORA ARNESI

INSTITUTO DE INVESTIGACIONES TEÓRICAS Y APLICADAS DE LA ESCUELA DE ESTADÍSTICA.

narnesi@fcecon.unr.edu.ar

La evaluación económica en salud abarca diversas técnicas o procedimientos para recabar información sobre la relación que existe entre el costo y el resultado de implementar determinadas intervenciones en salud. Si la medida de resultado que se incorpora al proceso de toma de decisiones se relaciona con la experiencia del paciente ante el evento clínico de interés, la evaluación se denomina: Análisis de Costo-Efectividad. La medida de mayor interés en estas evaluaciones económicas en salud es la Razón de Costo-Efectividad Incremental conocida habitualmente por su sigla en inglés ICER, la cual se define como el cociente entre el incremento del valor esperado de los costos de las intervenciones a comparar y el incremento del valor esperado de las efectividades correspondientes. Resulta natural intentar cuantificar la incertidumbre en la estimación del ICER a través de la construcción de intervalos de confianza, sin embargo uno de los problemas que se presenta es la necesidad de realizar supuestos acerca de la distribución muestral del estimador del ICER sumado a complejidad de la obtención de la variancia del estimador de una razón. Para subsanar este inconveniente, recientemente, gran parte de la bibliografía sobre el tema se ha focalizado en el uso de métodos no paramétricos. De ahí que el objetivo de este trabajo es comparar los resultados hallados a partir de métodos derivados del enfoque paramétrico clásico y el correspondiente al método no paramétrico basado en Bootstrap. Se comparan los Intervalos de Confianza, calculados a partir de los datos obtenidos de la simulación de una muestra, bajo el esquema de un árbol de decisión de 200 pacientes, de una situación hipotética donde se comparan dos técnicas quirúrgicas. El valor estimado del ICER es aproximadamente igual a 33 y representa el costo de una unidad adicional de efectividad. El intervalo más estrecho se logra a partir de la utilización del método no paramétrico y resulta igual a [30.25 ; 35.47]. En cambio, los métodos tradicionalmente utilizados en el enfoque paramétrico arrojan intervalos de confianza menos precisos. El intervalo más amplio se obtiene por la aplicación del método de Fieller, [2.61 ; 160.02], seguido por el que utiliza la aproximación de la variancia por series de Taylor y el método Confidence Box. Los resultados hallados en este estudio por simulación, sumados a la ventaja de no requerir supuestos distribucionales apoyan la recomendación del uso del método Bootstrap para la determinación de los límites de confianza.

Palabras claves: economía en salud; razón costo-efectividad; estimación por inter

TOMA DE DECISIONES EN EL SECTOR SALUD: ANÁLISIS DEL TIEMPO DE ESTANCIA POR ACCIDENTES DE TRÁNSITO

MARÍA VIRGINIA PISANI¹, FERNANDA VILLARREAL² y EUGENIA ELORZA³

¹DEPARTAMENTO DE MATEMÁTICA- UNIVERSIDAD NACIONAL DEL SUR. ²DEPARTAMENTO DE MATEMÁTICA- UNIVERSIDAD NACIONAL DEL SUR; INSTITUTO DE INVESTIGACIONES ECONÓMICAS Y SOCIALES DEL SUR (CONICET-UNS). ³DEPARTAMENTO DE MATEMÁTICA- UNIVERSIDAD NACIONAL DEL SUR; INSTITUTO DE INVESTIGACIONES ECONÓMICAS Y SOCIALES DEL SUR (CONICET-UNS); DEPARTAMENTO DE ECONOMÍA- UNIVERSIDAD NACIONAL DEL SUR.

fvillarreal@uns.edu.ar

La variable duración de la estancia (length of stay (LOS)) representa el tiempo que un paciente

permanece internado en una institución hospitalaria y constituye un instrumento clave de gestión ya que determina un alto porcentaje del gasto hospitalario. En las internaciones asociadas a los procesos de atención de pacientes que fueron víctimas de un accidente de tránsito el comportamiento de la variable duración de la estancia brinda información que puede ser de utilidad para predecir la evolución clínica del paciente como para la toma de decisiones que mejoren la gestión de los recursos hospitalarios afectados por estos procesos de atención y por este motivo, es sumamente relevante avanzar en el análisis estadístico. El objetivo de este trabajo es estudiar la distribución del tiempo de internación de pacientes que ingresaron por accidentes de tránsito a una institución hospitalaria en la ciudad de Bahía Blanca. Se utilizó una muestra de 121 casos que llegaron a dos hospitales públicos (uno provincial y otro municipal) entre el 22 de noviembre de 2014 al 22 de mayo 2015. Se realizó en primer lugar un análisis descriptivo y posteriormente un análisis inferencial utilizando los contrastes no paramétricos de Mann-Whitney y Kruskal-Wallis para probar si existen diferencias significativas en la distribución de la variable de interés según las variables sociodemográficas (edad, género, medio de locomoción, localidad y hospital) que pueden afectar el tiempo de permanencia en el hospital. Entre los principales resultados hallados se puede mencionar que: 1) la distribución del tiempo de internación presenta asimetría positiva en ambos hospitales; 2) existen diferencias significativas entre las distribuciones del tiempo de internación según el medio de locomoción utilizado por el accidentado, observándose que la permanencia en el hospital fue mayor para los casos en donde la víctima se transportaba en un auto sin cinturón de seguridad y 3) existen diferencias significativas para la variable en estudio de acuerdo a si el accidente se produjo dentro de la ciudad o si el paciente llegó derivado de otra localidad de la región. Los resultados obtenidos en base a la metodología empleada, pueden ser relevantes tanto para la gestión hospitalaria como para el diseño de las políticas de seguridad vial, al mejorar la calidad de la información en base a la cual se toman las decisiones de asignación de recursos y se estiman los costos asociados a este problema de salud pública.

Palabras claves: estadística inferenc; tiempo de estancia; accidentes de tránsi; gestión hospitalaria

Métodos bayesianos

AJUSTE DE MODELOS POISSON Y BINOMIAL NEGATIVO TRUNCADOS EN CERO MEDIANTE UN ENFOQUE BAYESIANO PARA DATOS DE CONTEO DE MEDICAMENTOS UTILIZADOS POR ADULTOS MAYORES

LUCIANA CARLA CHIAPELLA y MARÍA EUGENIA MAMPRIN

ÁREA FARMACOLOGÍA, FACULTAD DE CIENCIAS BIOQUÍMICAS Y FARMACÉUTICAS, UNIVERSIDAD NACIONAL DE ROSARIO, CONICET.

lucianachiapella@conicet.gov.ar

Introducción: Los adultos mayores conforman una población caracterizada por el uso de un alto número de medicamentos de forma diaria y crónica. En los estudios de utilización de medicamentos, la cantidad de fármacos utilizados por paciente es una variable de conteo que podría analizarse mediante modelos Poisson. Sin embargo, el supuesto de equidispersión es muy restrictivo, más aún cuando se está en presencia de datos truncados. Para estos casos, se sugiere utilizar modelos de regresión considerando la distribución Binomial Negativa truncada para modelar la variable respuesta. **Objetivos:** Comparar los resultados obtenidos mediante el ajuste de modelos Poisson y

Binomial Negativo mediante técnicas bayesianas, considerando que los datos se encuentran truncados en cero. Evaluar el posible efecto del sexo y la edad de los pacientes sobre la cantidad de medicamentos utilizados en forma crónica y simultánea. Materiales y métodos: Se registraron los medicamentos dispensados a pacientes mayores de 65 años en 10 farmacias comunitarias de Rosario, entre los meses de abril y septiembre de 2015. Dicho periodo fue dividido en dos trimestres y se contabilizaron los medicamentos dispensados a cada paciente en ambos ciclos. Mediante un enfoque bayesiano, se ajustaron modelos Poisson y Binomial Negativo con distintas distribuciones a priori para el parámetro de dispersión (log-normal y uniforme). Para ello, se realizaron simulaciones mediante el método Monte Carlo Markov Chain (MCMC), utilizando el software Winbugs. Se evaluó la relación entre la edad y el sexo de los pacientes y la cantidad de medicamentos utilizados. Resultados: Según el Deviance Information Criterion (DIC), se verificó que el modelo Binomial Negativo truncado en cero ofrece mejor ajuste que aquel que considera la distribución Poisson para los datos de conteo, incluyendo la edad como variable explicativa (DIC 9315 versus 9761). El efecto del sexo de los pacientes sobre la cantidad de medicamentos utilizados resultó no significativo. La edad de los individuos presentó una relación estadísticamente significativa con la variable respuesta, evidenciando un crecimiento en el número de medicamentos empleados de acuerdo con el aumento en la edad. Conclusiones: Para este caso particular de análisis de datos de conteo positivos, se verificó la utilidad de los modelos binomiales negativos truncados para la modelización de los mismos. Como era previsible este modelo permitió establecer una relación significativa entre el número de medicamentos utilizados y la edad de los pacientes.

Palabras claves: binomial negativo; mcmc; datos de conteo; medicamentos

INCLUYENDO PROPIEDADES EMERGENTES AL ESTUDIO DE MOVIMIENTO ANIMAL: UN ENFOQUE ESPACIO ESTADO CON AJUSTE POR ABC

SOFIA RUIZ SUAREZ y JUAN MANUEL MORALES

INIBIOMA—CONICET, UNIVERSIDAD NACIONAL DEL COMAHUE, BARILOCHE, ARGENTINA.

sofi.ruizsuarez@gmail.com

La forma en que se mueven los organismos es un proceso de fundamental importancia en ecología. El mismo determina el destino de los individuos en el sentido que interfiere en su adecuación biológica, en la evolución de las especies, en la estructura de las poblaciones y comunidades, y en la transmisión de enfermedades. En los últimos años, a partir de los nuevos avances en tecnología de rastreo y monitoreo, distintos métodos de estadística computacional han permitido ajustar modelos de movimiento cada vez más complejos. Sin embargo, cuando se simulan modelos ajustados a datos, las trayectorias resultantes raramente se parecen a las observadas. Una posible causa de este problema puede ser que estos modelos basan el ajuste solo en propiedades de las trayectorias a pequeña escala como son el largo de los pasos o los ángulos de giro, sin considerar propiedades emergentes del comportamiento de los animales en el paisaje como el uso del espacio. De forma de contemplar este punto, en este trabajo se presenta un modelo de movimiento basado en caminatas aleatorias dentro de un marco espacio estado, cuyo ajuste se plantea a partir de técnicas de Approximate Bayesian Computation (ABC) en donde los valores resumen contemplados en el ajuste, buscan reflejar propiedades emergentes del movimiento. Con este propósito, se consideraron estadísticos resumen tales como la medida de tortuosidad de la trayectoria, la función de auto correlación entre el largo de los pasos, el índice de sinuosidad, así como también se incluyeron la media del largo de paso y de los ángulos de giro, entre otros. Como medida de distancia para la comparación entre las simulaciones y las observaciones se consideró la distancia de Mahalanobis. Los modelos espacio-estado permiten enmarcar este análisis en un contexto donde es posible contemplar tanto el proceso de movimiento del individuo como el proceso de observación. A partir de datos simulados

se exploraron distintas posibilidades bajo diferentes escenarios, de forma de destacar las ventajas y limitaciones de este enfoque. Se encontró que dependiendo en parte de la escala de observación y del tipo de movimiento puede resultar beneficioso incorporar este tipo de información al ajuste. Consideramos útil seguir trabajando en esta línea de investigación, contemplando modelos de movimiento con más estructura que permitan mejores distinciones entre trayectorias.

Palabras claves: approximate bayesian; modelos espacio esta; movimiento animal

NUEVO ENFOQUE PARA LA BÚSQUEDA DE GENES ISOFUNCIONALES Y SU APLICACIÓN EN EL ESTUDIO EN *BACILLUS* PROMOTORES DEL CRECIMIENTO VEGETAL

MARIANO TORRES MANNO¹, MARÍA DOLORES PIZARRO², PAMELA BELÉN BARENGO³, LUCAS DAURELIO² y MARTÍN ESPARIZ¹

¹IBR-CONICET. ²CÁTEDRA DE FISIOLOGÍA VEGETAL, FACULTAD DE CIENCIAS AGRARIAS, UNL; CONICET. ³CÁTEDRA DE FISIOLOGÍA VEGETAL, FACULTAD DE CIENCIAS AGRARIAS - UNL.

mariano.atm@gmail.com

Las rizobacterias promotoras del crecimiento de plantas (PGPR por “Plant Growth Promoting Rhizobacteria”) son bacterias beneficiosas que tienen la capacidad de colonizar las raíces de las plantas y promover su crecimiento. Los Bacillus son preferidos a la hora de formular productos comerciales dado que al ser formadores de esporas tienen mayor viabilidad en el tiempo. Las vías por las cuales las bacterias promueven de diferentes formas el crecimiento en las plantas son diversas. Una de las más importantes es la síntesis de metabolitos secundarios (SMS) que incluyen polímeros como los policétidos y pequeñas estructuras tales como bacteriocinas, oligopéptidos y lipopéptidos. Dada la complejidad de secuencia de los genes que codifican para estas vías su asignación suele ser errónea. Aunque existen buenos algoritmos para la identificación de clusters génicos de SMS, dada la similitud de secuencia y naturaleza modular de los genes que componen cada cluster, la predicción de isofuncionalidad entre clusters codificados en distintas cepas sigue representando un verdadero desafío. Por ello se ha desarrollado un flujo de trabajo, basado en scripts ad hoc de R que utilizan BLAST, un enfoque inferencial Bayesiano y análisis de sintenia, para estimar la probabilidad de que una dada vía de SMS sea isofuncional a aquellas usadas como carnada. Se ha realizado la búsqueda y caracterización de 67 genes de las 10 vías SMS más importantes en 729 *Bacillus* con genoma no redundante disponibles al momento. Por medio de alineamientos por BLAST identificamos 27.038 genes que compondrían 6.834 vías SMS. Utilizando histogramas de los porcentajes de identidad de poblaciones de ortólogos y parálogos se infirió densidad a priori. Con las curvas de densidad se obtuvo tanto el factor bayesiano como la probabilidad de ortólogo o parálogo de cada alineamiento de BLAST y se determinó que solo 17.854 genes serían isofuncionales a los usados como carnada. Estos conforman solo 4.299 vías lo que supone que la herramienta BLAST genera un 40% de falsos positivos en este tipo de análisis. Analizando la sintenia de todas las 4.299 vías utilizando un script de R ad hoc pudimos determinar que 2.312 presentaban una estructura similar a las de las vía query. Los pipelines y scripts generados pueden ser utilizados para múltiples propósitos que involucren búsqueda de genes con actividades deseadas en diferentes genomas en forma automática y altamente eficiente lo que permitirá profundizar y sistematizar estudios de genómica comparativa.

Palabras claves: pgpr; bacillus; bayes; genómica comparativa

REGRESIÓN EN GRANDES DIMENSIONES. ESTUDIO COMPARATIVO DE LA REGRESIÓN LASSO BAYESIANA Y UNA ALTERNATIVA LASSO BAYESIANA ROBUSTA

MARÍA BELÉN ALLASIA¹, MARCIA D'ELIA BRANCO² y MARTA BEATRIZ QUAGLINO³

¹IITAE - CONICET. FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA. UNIVERSIDAD NACIONAL DE ROSARIO. ²INSTITUTO DE MATEMÁTICA E ESTADÍSTICA, UNIVERSIDADE DE SÃO PAULO. ³IITAE. FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA. UNIVERSIDAD NACIONAL DE ROSARIO..

mallasia@fcecon.unr.edu.ar

En el análisis estadístico de grandes volúmenes de datos es indispensable identificar aquellas variables que provean información valiosa. Al ajustar modelos de regresión, el método tradicional de estimación brinda estimadores insesgados pero poco precisos al trabajar con gran número de predictores, por lo que se proponen métodos de regularización para mejorar estos ajustes. Estos métodos, también conocidos como penalizados, consisten en ajustar el modelo eliminando predictores en el proceso de estimación, forzando que los coeficientes pequeños sean igual a cero. Entre éstos, la regresión LASSO (Least Absolute Shrinkage and Selection Operator) se ha convertido en una alternativa ampliamente utilizada. Sin embargo, la aplicación de técnicas clásicas requiere el cumplimiento de supuestos que rara vez son verificados en la práctica, por lo que varios autores han trabajado en el desarrollo de estimadores robustos, probando que éstos tienen un desempeño similar a los clásicos cuando se cumplen los supuestos. En el presente trabajo se estudia la metodología LASSO implementada mediante el enfoque estadístico Bayesiano, para la estimación de los parámetros de regresión en el contexto de muchas variables predictoras. Se propone una versión de Regresión LASSO Bayesiana Robusta considerando en el modelo jerárquico que los errores aleatorios siguen una distribución a priori t-Student, lo que se consigue asumiendo que los errores aleatorios siguen una distribución normal con mezcla de escala, cuya variancia tiene una distribución hiperpriori Gamma; en lugar del supuesto tradicional de distribución normal. Se simularon diferentes escenarios con mayor número de parámetros que observaciones, cambiando el número de parámetros significativos a estimar. Se comparó el desempeño de ambos métodos a través del Error Cuadrático Medio (ECM). Siguiendo el principio propuesto por Hastie, Tibshirani y Friedman (2009) de “apostar al esparcimiento” -que propone elegir aquel procedimiento que funcione bien en problemas de datos esparcidos, ya que ningún método funciona bien en problemas densos-, se generaron escenarios donde el número de parámetros significativos fuera pequeño en relación al total de variables predictoras. Al incrementar el número de parámetros significativos, el ECM de cada técnica aumenta, siendo estrictamente menor el ECM de la Regresión LASSO Bayesiana Robusta y con menor variabilidad en todos los casos. Además, al asumir un modelo no normal para las observaciones, la regresión penalizada robusta otorga mejores estimaciones, siendo más precisos los estimadores calculados. En conclusión, se recomienda usar estimaciones robustas de los parámetros de regresión ya que proveen un ajuste más apropiado del comportamiento de los datos en cualquier contexto, sin demandar el cumplimiento de supuestos acerca del comportamiento de los mismos.

Palabras claves: data mining; regresión penalizada; regresión robusta; selección de variabl

Métodos multivariados

ALGORITMOS DE AGRUPAMIENTO MULTIVARIADOS PARA GRANDES BASES DE DATOS

MARÍA EUGENIA VIDELA, JULIANA IGLESIAS y CECILIA BRUNO

eugeniamidelal2@gmail.com

La estructura genética poblacional entre genotipos permite conocer relaciones entre los individuos e incorporar dicha información en modelos de asociación y/o predicción genómica. Sin embargo, aplicar métodos de agrupamiento en una gran colección de datos implica un alto costo computacional y un incremento en la complejidad para la búsqueda de estructura genética. Diferentes algoritmos de agrupamiento han surgido como una alternativa para analizar con precisión el volumen masivo de datos generados por las aplicaciones modernas. El objetivo de este trabajo fue evaluar algoritmos de agrupamiento en grandes bases de datos genéticas obtenidas a partir de marcadores moleculares del tipo Single Nucleotide Polymorphism (SNP). En el presente trabajo se ilustra la comparación de diez algoritmos de clasificación: Divisive Analysis Clustering (DIANA), Partitioning Around Medoids (PAM), Clustering Large Applications (CLARA), Agglomerative Nesting (AGNES), Unweighted Pair Group Method with Arithmetic Mean (UPGMA), K-means, Kernel K-means, Fuzzy K-means, Admixture y Self Organizing Tree Algorithm (SOTA), sobre una base de datos de 270 líneas estabilizadas de maíz genotipadas con 55769 marcadores moleculares tipo SNP pertenecientes al Grupo de Mejoramiento de maíz de INTA Pergamino. Los marcadores fueron codificados en forma binaria (donde 1 representa alelos heterocigotas y 0 alelos homocigotas), se eliminaron aquellos marcadores con frecuencia alélica menor a 0.01 y aquellos con más del 30% de datos faltantes, obteniendo así un total de 51576 marcadores moleculares. La validación del número óptimo de grupos para cada algoritmo comparado se realizó mediante los índices de Dunn, Conectividad y Silueta. Todos los métodos sugirieron la existencia de dos subpoblaciones, excepto los métodos basados en K-means cuyos índices de Silueta indican un número mayor de grupos. De la comparación entre métodos se concluyó que DIANA, AGNES y UPGMA presentaron mejor desempeño en la clasificación de la base de datos genética.

Palabras claves: estructura genética; marcadores snp; clasificación multiva

ANÁLISIS DE COMPONENTES PRINCIPALES CON DATOS GEORREFERENCIADOS. UNA APLICACIÓN EN LA INDUSTRIA TURÍSTICA

LAURA ISABEL LUNA y FERNANDO GARCÍA

FCE, UNC.

lauraisabel.luna@gmail.com

Análisis de componentes principales con datos georreferenciados. Una aplicación en la industria turística. Luna, Laura Isabel*, García, Fernando* Facultad de Ciencia Económicas, UNC lauraisabel.luna@gmail.com, fgarcia.unc@gmail.com Existen nuevos métodos multivariados que permiten mapear el espacio geográfico según la estructura espacial de indicadores económicos como son los datos de actividades relacionadas al turismo que conforman el producto bruto. Dray et al. (2008), proponen un método de análisis multivariado que incorpora la información espacial previo al

análisis multivariado, el método es conocido como MULTISPATI-PCA. La técnica ha mostrado ser eficiente en estudios de agricultura de precisión (Córdoba, 2014), y en este trabajo se prueba para variables económicas. Existen interacciones subyacentes entre las diferentes actividades económicas, por ello el análisis de las covariaciones o correlaciones entre las mismas es un aspecto que debe ser considerado en estudios económicos de naturaleza multivariada. No obstante, es importante remarcar que la estructura de covariación reflejada por un análisis multivariado clásico puede verse afectada por los patrones espaciales subyacentes en los datos. Las componentes principales (PC) son apropiadas sólo para resumir variabilidad y no están diseñadas para revelar patrones espaciales. Por ello es necesario utilizar una metodología que resuma la variabilidad y revele estructuras espaciales al mismo tiempo; existen hoy métodos que abarcan estos dos objetivos. En este trabajo se exponen dos técnicas y se realiza un análisis comparativo de los resultados obtenidos, con la implementación de un PCA clásico y de una versión restringida espacialmente (MULTISPATI-PCA, Dray et al., 2008). Se utilizaron datos procesados por la Dirección de Estadísticas Económicas (DEE) de la provincia de Córdoba para el periodo 2001 – 2014. Los mapas de variabilidad espacial construidos a partir de la primera componente de ambas técnicas fueron similares; no así los de la segunda componente debido a cambios en la estructura de co-variació identificada, al corregir la variabilidad por la autocorrelación espacial de los datos. Palabras Clave: análisis multivariado, multispati-pca, pca, actividades económicas .

Palabras claves: análisis multivariado; multispati-pca; pca; actividades económicas

ANÁLISIS DEL CUMPLIMIENTO DE LOS SUPUESTOS DEL MODELO AMMI Y SU INFLUENCIA EN LA SELECCIÓN DE GENOTIPOS DE DURAZNERO

*JULIA ANGELINI¹, LUIS ARROYO², MARÍA ELENA DAORDEN², GABRIEL
HUGO VALENTINI², MARTA BEATRIZ QUAGLINO³ y GERARDO DOMINGO
LUCIO CERVIGNI¹*

¹CEFOBI-CONICET. ²INTA SAN PEDRO. ³ESCUELA DE ESTADÍSTICA - UNR.

angelini@cefobi-conicet.gov.ar

La identificación de genotipos superiores en diferentes ambientes es el principal objetivo del mejoramiento genético. En duraznero, la interacción genotipo x ambiente (IGA) del tipo compleja, altera el desempeño de genotipos haciendo que la selección sea una tarea compleja. El modelo AMMI (Additive Main effects and Multiplicative Interactions), que combina el análisis de la variancia (ANOVA) de efectos genotipo y ambiente, seguido de un análisis de componentes principales (ACP) de los residuos resultantes, ha ganado importancia en ensayos multiambientales. La validez del modelo depende del cumplimiento de dos supuestos, que los errores posean distribución normal y que la varianza de los mismos, entre campañas, sean homogéneas, los que raramente se evalúan o verifican en datos agronómicos. En este trabajo se analizó la robustez de AMMI y del parámetro ASV (AMMI Stability Value) y GSI (Genotype Selection Index), derivados del AMMI, para la selección del 20% de los genotipos evaluados. La eficiencia obtenida fue comparada con la de los índices de superioridad (P_i) y de confianza (I_i) los cuales no son derivados de AMMI y cuyo buen desempeño en duraznero fue demostrado previamente. Se analizaron datos de rendimiento (kg/planta) de 29 genotipos de duraznero, obtenidos de siete campañas en INTA San Pedro. Como los supuestos de AMMI no se cumplieron, se utilizó la transformación indicada por el método de Box-Cox, potencia de 0,28. En el ANOVA inicial la IGA explicó el 63,63% de la variación de los efectos ($P<0,0001$), aunque los supuestos no se cumplieron debido a la presencia de valores atípicos. Cuando en el ANOVA se incluyó solamente ambiente y genotipo, la transformación logró una distribución normal de los errores con variancias constantes. Las dos primeras componentes explicaron 57,20% y 55,86% de la variabilidad residual considerando la variable original y trasformada, re-

spectivamente. Al transformar la variable, ASV coincidió en la selección de 1 de los 6 genotipos seleccionados con la variable original; mientras que con GSI se seleccionaron 4 de 6. Los métodos Pi e Ii seleccionaron los mismos seis genotipos con ambos tipos de variables. Cinco de los seleccionados con Pi e Ii coincidieron con GSI y dos con ASV usando la variable transformada, y de cuatro y uno para GSI y ASV, respectivamente, cuando se usó la variable original. Por lo tanto, al transformar la variable, se mejora levemente la eficiencia de selección de los parámetros derivados de AMMI.

Palabras claves: adaptabilidad; estabilidad; métodos multivariado; multiambiente; prunus persica

ANÁLISIS ESTADÍSTICO DE LOS PATRONES DE COMPORTAMIENTO DE LA SUCCIÓN NUTRITIVA EN RECIÉN NACIDOS A TÉRMINO Y PREMATUROS

PATRICIA GIRIMONTE¹, LAURA ALVAREZ², NATALIA ELISEI², SILVIA JURY³ y ANIBAL PEDRO LAQUIDARA⁴

¹UBA - CÁTEDRA DE MATEMÁTICA, FFYB, Y DEPARTAMENTO PEDAGÓGICO DE MATEMÁTICA, FCE. ²UBA - FACULTAD DE MEDICINA.
³IDIP. ⁴UNLP, CIOP-CICBA.

patriciagirimonte@gmail.com

La cuantificación y caracterización de los patrones de succión nutritiva (SN) resultan interesantes y pertinentes clínicamente; siendo esta una actividad que puede ser fácilmente observable cuando se alimenta a un bebé. En Argentina existen antecedentes de evaluación de recién nacidos a término (RNT) pero hasta el momento no en recién nacidos pre término (RNPT). El objetivo del presente trabajo es analizar los patrones de comportamiento de la SN en RNPT y su comparación con los RNT. El estudio consta de una muestra que incluyó 45 RNT sin patología crónica asociada y 75 RNPT concurrentes al Hospital Sor María Ludovica de la ciudad de La Plata. Los recién nacidos fueron evaluados en condiciones similares, con posición de 45°, en su toma habitual (mamadera estándar y leche que bebía diariamente). Se utilizó un equipo electrónico de manometría (transductor de presión y conexión a computadora) que permite observar y registrar la amplitud [mmHg] y la frecuencia [Succ/seg] de succión. Luego se procesaron las señales con un algoritmo semi-automático. Las mediciones las realizó un solo evaluador en la zona de regularidad de la curva. Se analizó la relación entre las variables amplitud [mmHg] y frecuencia [Succ/seg] de succión. Para el análisis se aplicó la transformación logarítmica a ambas variables. Además de un análisis descriptivo dentro de cada grupo, se realizaron elipses de confianza de nivel 0.95 basadas en la distancia de Mahalanobis; se validó el supuesto de normalidad con el test multivariado de Shapiro Wilks y los test de Mardia. Se aplicó el test T2 de Hotelling resultando la diferencia entre los vectores de medias de ambos grupos estadísticamente significativa. Dentro de cada grupo se analizó la correlación entre amplitud y frecuencia, resultando negativo el coeficiente de correlación de Pearson, pero solo significativo en los RNT. Estos resultados sugieren que existirían diferentes aspectos del patrón de succión que maduran diferencialmente. Estos hallazgos coinciden con lo referido por otros autores interpretando que el comportamiento del prematuro muestra cierta desincronización e inmadurez con respecto al RNT. El análisis estadístico se realizó con R 3.4.0.

Palabras claves: succión nutritiva; recién nacidos pre t; análisis multivariad; correlación

ANALISIS ESTADISTICO DE PUNTAS LITICAS Y SU CORRESPONDENCIA CON DIFERENTES SISTEMAS DE ARMAS

ALICIA HERNÁNDEZ¹, LILIANA A GARCÍA², RODRIGO VECCHI³ y CRISTINA BAYÓN³

¹UTN FRBB - UNS. ²UNS. ³CONICET - UNS.

aliciahe@criba.edu.ar

Las puntas líticas son el remanente perdurable de un artefacto conformado por varias tecnologías, que sólo puede estudiarse de manera indiciaria. Para el abordaje de las puntas líticas, un buen punto de partida consiste en la realización de estudios tecno-morfológicos de artefactos procedentes de microrregiones y relacionar el análisis de los diseños con las asociaciones contextuales y cronológicas. La descripción y clasificación de las puntas líticas permiten reconocer ciertas tendencias o patrones generales. En la pampa bonaerense, las puntas líticas se han hallado en distintos sitios, desde las primeras ocupaciones (ca. 12.000 años AP) hasta momentos previos al contacto con los europeos, con características tecno-morfológicas variadas. La presencia de estos diferentes tipos de puntas líticas generó interés sobre los sistemas de armas y sus componentes y sobre las características tecno-tipológicas. El objetivo de este trabajo es analizar las puntas de proyectil (78 piezas) provenientes de ocho sitios de dicho sector a fin de caracterizarlas y determinar si provienen de diferentes sistemas de armas. En primer término, se registraron las variables métricas y en base a éstas se realizó la segmentación del conjunto en grupos mediante análisis estadístico, descriptivo e inferencial, aplicando métodos univariados y multivariados. Del análisis descriptivo de 78 piezas pudo observarse una variabilidad relativa superior al 10 % para las variables longitud, ancho y espesor. Dado la heterogeneidad de las piezas, mediante análisis multivariado se las segmentó en cuatro grupos. Para determinar si los grupos presentan diferencias estadísticamente significativas para las variables métricas en estudio, se realizaron pruebas de comparación de medias, paramétricas o no paramétricas, según se cumpla o no el requisito de homocedasticidad. Las mismas resultaron altamente significativas (valor $p < 0,01$), lo que permitió concluir que, para cada variable, al menos uno de los grupos difería de los demás. Posteriormente, se realizó la prueba T3 de Dunnet de comparaciones múltiples, para cada variable, concluyendo que todas las medias difieren entre sí (valor $p < 0,01$). Es decir que los tres grupos presentaron diferencias estadísticamente significativas. El análisis estadístico fue realizado mediante el software Infostat versión 2013. Como conclusión podría plantearse, inicialmente, que las diferencias entre los grupos estarían relacionadas con la presencia de diferentes sistemas de armas.

Palabras claves: análisis multivariado; anova; comparaciones múltiples; puntas líticas; sistemas de armas; características técnicas

APLICACIÓN DE METODOLOGÍA NO PARAMÉTRICA Y MULTIVARIADA PARA EL ANÁLISIS DE ATRIBUTOS DE UN CONJUNTO DE YERBAS COMERCIALES

MYRIAM NUÑEZ¹, MARCELA FERNÍCOLA¹, FABIÁN DRUNDAY² y AMALIA CALVIÑO³

¹UNIVERSIDAD DE BUENOS AIRES, FACULTAD DE FARMACIA Y BIOQUÍMICA, CÁTEDRA DE MATEMÁTICA. ²UNIVERSIDAD DE BUENOS AIRES, FACULTAD DE FARMACIA Y BIOQUÍMICA, CÁTEDRA DE FISIOLOGÍA. ³IQUIMEFA, UBA-CONICET.

myriam@ffyb.uba.ar

La yerba mate presenta un elevado índice de consumo doméstico y también se diversifica su em-

pleo en numerosos mercados a nivel mundial (Braceco y col., 2010, Filip, 2011). En los últimos años el sector yerbatero ofrece productos diferenciados, con mayor variedad de sabores y amplios beneficios para la salud que incluye un porcentaje de numerosas hierbas agregadas dando lugar a la categoría de yerba mate compuesta. La diversidad de productos ofrecidos por el mercado hace relevante la descripción y el estudio de los mismos basándonos en la percepción de los consumidores, más representativos que los paneles entrenados. Los objetivos de este trabajo consisten en: a- analizar 5 tipos de yerba mate elaborada y compuesta al ser consumida por un grupo de 85 panelistas no entrenados de la zona metropolitana mediante un listado que incluye atributos sensoriales, emocionales y sociales, b- posicionar y agrupar los productos en función de las características mencionadas, c- dar algunos descriptores del producto considerado ideal y su ubicación relativa dentro del grupo analizado. El estudio se lleva a cabo a través de la metodología CATA (check-all-that-apply) y que consistió, en este caso, en una encuesta de 24 términos asociados al producto que los consumidores consideran relevantes en la elección de las yerbas. Como primera medida se aplicó la prueba Q de Cochran sobre cada atributo. Se obtuvieron así descriptores significativos en la diferenciación de las yerbas, entre los que se destacan los sensoriales (“nota a menta”, “peperina”, “refrescante”) y los emocionales/sociales (“calma tensiones”, “para tomar al aire libre”), etc. Luego se realizaron pruebas de a pares para identificar las muestras significativamente diferentes. A través Análisis de Correspondencias se obtuvo un mapa sensorial de las yerbas y términos de CATA que permite visualizar las similitudes y diferencias entre las muestras así como sus principales características sensoriales, determinando así la posición relativa de las mismas. Se identificaron tres grupos, el primero relacionado con atributos como “nota a hierbas”, “poco amargo” y “calma tensiones”, otro asociado a “mucho polvo”, “muy amargo” y “sabor persistente al tragarse”, y por último una yerba con “nota ahumada” y “sin polvo”. Finalmente el estudio reúne información para identificar los principales atributos sensoriales percibidos que nos permitan acercarnos al producto ideal.

Palabras claves: cata; atributos sensoriale; análisis de correspo; pruebas no paramétricas

APLICACIÓN DEL AGLORITMO BOOSTING ADAPTATIVO (ADABOOST) A UN PROBLEMA DE CLASIFICACIÓN AUTOMÁTICA DE TEXTOS

CELINA BELTRÁN y IVANA BARBONA

CÁTEDRA DE ESTADÍSTICA, FACULTAD DE CIENCIAS AGRARIAS, UNIVERSIDAD NACIONAL DE ROSARIO.

ivanabarbona@gmail.com

Boosting es un método que pretende mejorar el desempeño de cualquier algoritmo de aprendizaje supervisado mediante la combinación de los resultados de varios clasificadores débiles o de base para obtener un clasificador final robusto. Una de las técnicas más populares de Boosting es el algoritmo Boosting Adaptativo (AdaBoost). Este algoritmo, mediante un entrenamiento iterativo de los clasificadores débiles o de base, le asigna mayor importancia a los datos mal clasificados anteriormente, y de esta manera obtiene un nuevo clasificador. Logra, de esta forma, adaptarse y obtener mejores resultados aumentando la precisión del algoritmo. En el presente trabajo, con el objetivo de evaluar el desempeño del algoritmo AdaBoost, se aplican los métodos de clasificación Regresión Logística y SMO (Sequential minimal optimization), con y sin el algoritmo AdaBoost a un conjunto de textos. Luego, se comparan los resultados obtenidos de los métodos de clasificación al considerarse solos, con los resultados al considerarlos como algoritmo de base para AdaBoost. El criterio de clasificación utilizado fue el género al que pertenece el texto (Científico / No Científico). La caracterización de los textos está basada en la distribución de frecuencias de las categorías morfo-sintácticas. Los resultados finales de los distintos clasificadores considerados se evalúan mediante porcentajes de mala clasificación. Se observó que al aplicar AdaBoost teniendo en cuenta como algoritmo de base el método de Regresión Logística no se presentó una reducción en el porcentaje de

mala clasificación. En cambio, para el caso del método SMO como algoritmo de base, el porcentaje de mala clasificación bajó un 8,67%.

Palabras claves: métodos de clasifica; aprendizaje de máqui; adaboost; boosting; regresión logística; smo

APLICACIÓN DEL ANÁLISIS DE COMPONENTES PRINCIPALES PARA LA OBTENCIÓN DE PATRONES DE CONSUMO ALIMENTARIO EN UN ESTUDIO DE CASOS Y CONTROLES DE ENFERMEDAD COLELITIÁSICA

GUILLERMINA ISERN¹, AGUSTINA BERTOLA COMPAGNUCCI² y STELLA MARIS PEZZOTTO²

¹FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA. UNR. ²IDICER-CONICET. FACULTAD DE CIENCIAS MÉDICAS. UNR.

gisern@fcecon.unr.edu.ar

El análisis de componentes principales se ha convertido en una herramienta fundamental de la epidemiología nutricional para la obtención de patrones de consumo alimentario, ya que los mismos describen tendencias de ingestas características de diversos grupos. El objetivo de este trabajo fue determinar el patrón de consumo alimentario de los casos y de los controles en un estudio de identificación de factores de riesgo alimentarios de enfermedad colelitiasica (EC). Los casos son 51 pacientes con diagnóstico de EC, algunos de los cuales ya han sido colecistectomizados. Los controles son 69 personas a las que se les realizó una ecografía abdominal para descartar la presencia de cálculos asintomáticos. A todos los participantes se les realizó una entrevista personal y se aplicó un cuestionario de frecuencia de consumo (FFQ) con 210 ítems para la recolección de la ingesta alimentaria. Para establecer el tamaño de las porciones alimentarias se utilizó un atlas asociado al FFQ. Con los datos obtenidos se calcularon los gramos consumidos de los diferentes grupos de alimentos, generando 29 variables, que fueron incluidas en un análisis de componentes principales (ACP). De la aplicación del ACP se seleccionaron las 3 primeras componentes, que representan un 34% de la variabilidad total. Entre las tres primeras componentes principales obtenidas, se determinaron dos que permiten diferenciar los casos de los controles: la primera y la tercera. Los casos se caracterizaron por consumir mayor cantidad de grasas animales, vísceras, azúcar, bebidas azucaradas, papas, cereales pastas y granos, fiambres y embutidos, pollo con piel y carne vacuna grasa, además de consumir poca cantidad de vegetales rojos y amarillos, coles, otras frutas, y pescado. En cambio, los controles se caracterizaron, en su mayoría, por condiciones contrarias a las descriptas para los casos. Además, según la tercera componente, los controles se caracterizan por consumir mayor cantidad de pollo sin piel, frutas secas, carne vacuna magra, lácteos enteros y frutas frescas y bajas cantidades de pollo con piel. Así, quedaron conformados 2 patrones de consumo alimentario. De esta manera queda evidenciado que el análisis de componentes principales es una herramienta útil y acorde para la generación de patrones de consumo alimentario en vista a su aplicabilidad en mensajes de prevención de las enfermedades.

Palabras claves: análisis de componen; patrones de consumo; casos y controles; enfermedad coleliti

ASOCIACIÓN ENTRE VARIABILIDAD MOLECULAR Y FENOTÍPICA EN TOMATE

*A P DEL MEDICO¹, V G CABODEVILA¹, M S VITELLESCHI², A LAVALLE³ y
G R PRATTA¹*

¹IICAR (INSTITUTO DE INVESTIGACIONES EN CIENCIAS AGRARIAS DE ROSARIO), CONICET/UNR. ²IITAE (INSTITUTO DE INVESTIGACIONES TEÓRICAS Y APLICADAS DE LA ESCUELA DE ESTADÍSTICA), CIUNR/UNR. ³DEPARTAMENTO DE ESTADÍSTICA, UNIVERSIDAD NACIONAL DEL COMAHUE.

delmedico@iicar-conicet.gob.ar

Actualmente, en los programas de mejoramiento genético resulta de interés conocer las asociaciones entre la variabilidad genética y fenotípica para caracteres de interés agronómico. Existen diferentes tipos de marcadores moleculares que pueden aplicarse para estimar la variabilidad genética. Entre ellos, la técnica de AFLP (Amplified Fragment Length Polymorphism o Polimorfismo en la Longitud de los Fragmentos Amplificados) no requiere un conocimiento previo de la secuencia de ADN y brinda marcadores de tipo dominante y multiloci. Por otro lado, los SSR (Single Sequence Repeats o Repeticiones de Secuencia Simple, también llamadas microsatélites) requieren un conocimiento previo de la secuencia de ADN, siendo marcadores codominantes y de único locus. El objetivo de este trabajo fue comparar, mediante Análisis Factorial Múltiple (AFM), la utilidad de ambos tipos de marcadores para caracterizar la variabilidad molecular y estimar su asociación con la variación fenotípica en una generación segregante de tomate. Se evaluaron 66 individuos provenientes de la F2 del híbrido de segundo ciclo F1 ToUNR18xToUNR1, cuyos perfiles de AFLP y SSR fueron obtenidos por técnicas estándares. Los caracteres fenotípicos medidos fueron peso, diámetro, altura, forma, vida poscosecha, porcentaje de reflectancia, firmeza, coeficiente a/b, sólidos solubles, pH y acidez titulable. Se aplicaron dos AFM para relacionar en forma independiente cada tipo de marcador molecular que segregó en la forma mendeliana esperada según su comportamiento genético, con los caracteres fenotípicos bajo estudio. 29 marcadores AFLP y 17 marcadores SSR segregaron en forma mendeliana. En el AFM, la proporción de variación explicada por los dos primeros ejes principales fue levemente superior para AFLP (30,9% vs. 25,0% para SSR). Si bien la representación de las correlaciones entre las variables cuantitativas resultó similar en ambos análisis, se observaron algunas discrepancias que podrían estar explicadas por su diferente asociación con la variación molecular. Además, el agrupamiento de los individuos de acuerdo a la caracterización fenotípica y molecular fue diferente, lo que se correspondería con el diferente tipo de variación en el ADN que mide cada marcador. La asociación global, medida por el coeficiente RV, fue mayor para SSR que para AFLP (0,22 y 0,13, respectivamente). Se concluye que, a pesar de que los dos tipos de marcadores considerados brindan información diferente respecto a la variación en el ADN, ambos son útiles para caracterizar la variabilidad molecular y estimar su asociación con la variación fenotípica en la generación segregante de tomate analizada.

Palabras claves: análisis factorial m; mejoramiento genétic; afp; ssr

COMPETENCIAS DIGITALES PARA LA EDUCACIÓN SUPERIOR A DISTANCIA ADAPTACIÓN Y APLICACIÓN DEL CUESTIONARIO ACUTIC AL CASO DE LA FCE-UNC

ADRIÁN MAXIMILIANO MONETA PIZARRO, LAURA MONTERO, MARÍA ALEJANDRA JUÁREZ, MARIANA LASPINA, JOSEFINA DEPETRIS, BRUNO FAGNOLA y FABRIZIO SOFFIETTI

UNIVERSIDAD NACIONAL DE CÓRDOBA.

amoneta@eco.uncor.edu

La Facultad de Ciencias Económicas (FCE) de la Universidad Nacional de Córdoba (UNC) ofrece desde 2002 un Ciclo Básico a Distancia (CBD) para sus carreras de grado. Esta propuesta consiste en disponer, para cada asignatura del CBD, de una comisión cuyo cursado se realiza a distancia con el apoyo de una plataforma virtual. El objetivo de este CBD es reforzar estrategias encaminadas a disminuir problemas y dificultades con el acceso, la deserción, el retraso y el fracaso académico en los primeros años de cursado (Moneta Pizarro et al., 2017). Sin embargo, estudios exploratorios previos muestran que menos del 40% de los alumnos que optan por esta modalidad logran la regularidad y que entre las posibles causas de este fenómeno se encontrarían problemas relacionados con la preparación, uso y actitud de los estudiantes para las tecnologías de la información y comunicación (Blanch et al., 2013 octubre). De acuerdo a la literatura, el dominio tecnológico y las competencias digitales de los alumnos son fundamentales para el éxito de la educación a distancia (Moore y Kearsley, 2011). Resulta entonces importante contar con escalas de medición de los constructos correspondientes y que éstas sean válidas para identificar la asociación con el desempeño académico de los estudiantes en esta modalidad. El objetivo de este trabajo es adaptar el cuestionario de actitud, conocimiento y uso de tecnologías de la información y la comunicación (ACUTIC) en Educación Superior propuesto por Mirete Ruiz, García-Sánchez y Hernández Pina (2015) y validarla estadísticamente sobre una muestra del curso de Microeconomía I del CBD de la FCE-UNC mediante técnicas de análisis factorial y modelación de ecuaciones estructurales. Los resultados demuestran consistencia interna de la escala total adaptada ($=0.9474$) y de cada una de las sub-escalas (actitud= 0.9393 , conocimiento= 0.9045 y uso= 0.8868). Las estimaciones del modelo de medida resultaron satisfactorias reportando mayoría de índices con bondad de ajuste aceptable (RMSEA= 0.069 ; CFI= 0.931 ; SRMR= 0.079). Esto evidencia que el instrumento propuesto es confiable y útil para obtener indicadores de actitud, conocimiento y uso de tecnologías de la información y la comunicación en cursos similares del CBD de la FCE-UNC y para ser aplicado en investigaciones que busquen identificar el impacto de estos constructos sobre otras variables relevantes.

Palabras claves: competencias digital; tecnologías de la in; educación superior a; ecuaciones es-
tructur

CONCIENCIA DE DAÑO DEL HÁBITO TABÁQUICO EN FAMILIARES DE PACIENTES CON CÁNCER

*MYRIAM NUÑEZ¹, FERNANDA RUSSO¹, JOSÉ MARÍA LASTIRI² y
GUADALUPE PALLOTTA²*

¹UBA. FACULTAD DE FARMACIA Y BIOQUÍMICA. CÁTEDRA DE MATEMÁTICA. ²HOSPITAL ITALIANO DE BUENOS AIRES.

myriam@ffyb.uba.ar

El objetivo del estudio es comparar los resultados obtenidos entre el Grupo de pacientes (GP) y el Grupo de familiares (GF) en pacientes con neoplasias asociadas y no asociadas al hábito tabáquico. Hay aspectos relacionados al consumo de tabaco que son fundamentales, especialmente si uno plantea la posibilidad de desarrollar estrategias contra esta adicción. Para este análisis se realizó una encuesta en el Hospital Italiano de Buenos Aires (HIBA) a 390 familiares de pacientes con diagnóstico de cáncer y a 350 pacientes. La misma fue efectuada tanto al paciente como al familiar por separado. Se registraron los siguientes datos, en ambos grupos; Sexo, Edad, Factores de riesgo: hábito de fumar, frecuencia de cigarrillos, tiempo que lleva fumando, Características del hábito, Conocimiento acerca de la enfermedad, Deseos de cesación, Conciencia de daño. Se observó que en el GF el consumo es levemente inferior al conocido en nuestro medio (29,3% fumadores; 15,6% ex fumadores y el 55,1% no fumadores), y, seguramente relacionado con el mayor número de mujeres en este grupo (71,1%). El 17 % afirmó que no cesaría de fumar, aunque esto ayudara a su familiar, lo que sugiere más claramente el desconocimiento y un profundo grado de adicción al tabaco. En el GF su conocimiento parece ser algo mayor y sin embargo el consumo sigue siendo semejante al del GP. La

aplicación de metodología estadística multivariada como, Análisis Factorial de Correspondencias, Regresión Logística y Árboles de Clasificación, nos permitió tener una idea del comportamiento de todos los factores de riesgo en su conjunto. Se concluyó que los mecanismos psicológicos previos, (específicamente, la falta de reconocimiento de una relación causal entre el tabaco y la patología que lo afecta), relacionados con la adicción, o reactivos a la situación de enfermedad, le impiden tanto al paciente como al familiar solucionar tal adicción.

Palabras claves: factores de riesgo; hábito tabáquico; conciencia de daño; métodos multivariado

DIMENSIONES DE LA POBREZA RELATIVA

ALFREDO BARONIO, FAVIO NICOLÁS D'ERCOLE y ANA MARÍA VIANCO

FACULTAD DE CIENCIAS ECONÓMICAS – UNIVERSIDAD NACIONAL DE RÍO CUARTO.

anavianco@yahoo.com.ar

La medición multidimensional de la pobreza incorpora aspectos no monetarios y evalúa si las personas logran alcanzar umbrales mínimos de bienestar en cada una de las dimensiones consideradas. En acuerdo a CEPAL (2013), identificar las dimensiones y establecer los umbrales son instancias que generan controversias; aunque mediciones en Colombia, México, Chile y Argentina coinciden en considerar educación, salud y nivel de vida. El relevamiento realizado por el Consejo Económico y Social en el año 2014 sobre 654 hogares, permitió conocer que en Río Cuarto la pobreza multidimensional relativa se asocia a nivel educativo que sólo alcanza a secundario incompleto, nivel de ingresos que sitúa a sus miembros en situación de riesgo alimentario y disponibilidad de bienes que no le permite acceder al confort derivado de las nuevas tecnologías; se consideró que el factor principal y discriminante de la pobreza está representado por la educación, siendo el umbral a superar el de la escuela secundaria. El estudio se repite en el año 2017 sobre 718 hogares, utilizando similar instrumento de recolección de información y recurriendo a estadística descriptiva, análisis factorial de correspondencias múltiples y clasificación jerárquica de la información para su análisis. El objetivo de este trabajo es comparar las modalidades predominantes que alcanza cada dimensión entre 2014 y 2017 y proponer los umbrales a considerar en cada una de ellas. Ambos estudios se realizaron en el segundo trimestre del año respectivo; a través de muestreo probabilístico de etapas múltiples, donde el radio censal tiene probabilidad 1 de formar parte de la muestra; en los hogares seleccionados se observaron las dimensiones ambiente, educación, empleo, salud, violencia, opinión e ingresos. Resultados preliminares identifican, nuevamente, a las dimensiones educación, empleo, ingresos y vivienda discriminando a la población, dando la posibilidad de hablar de pobreza relativa. A tres años del primer estudio, la situación social se ha mantenido estable con señales de mejoras en la dirección deseada, pero estas diferencias no resultan significativas.

Palabras claves: educación; umbral; diferencias; comparación

FACTORES DETERMINANTES DE LA ADOPCIÓN DE TICS EN EMPRESAS MANUFACTURERAS DE LA PROVINCIA DE CÓRDOBA

PABLO ORTIZ, MARIANA GUARDIOLA y NORMA P. CARO

FACULTAD DE CIENCIAS ECONÓMICAS, UNIVERSIDAD NACIONAL DE CÓRDOBA.

pabort@eco.uncor.edu

El propósito del presente trabajo es indagar sobre los distintos factores que inciden en la adopción de Tecnologías de la Información y la Comunicación (TICs) en empresas manufactureras de la

provincia de Córdoba. Dada la gran cantidad de variables, tanto de naturaleza cuantitativa como cualitativa, que influyen en este aspecto resulta menester su análisis conjunto, considerando las correlaciones que puedan existir y, eventualmente, las dimensiones que puedan constituir. A tal efecto, se utilizaron datos de la Encuesta de Innovación y Conducta Tecnológica del año 2014, diseñada e implementada por la Dirección General de Estadística y Censos de Córdoba, la cual cuenta con un módulo específico referido a las TICs. Particularmente, en 2014 fueron relevadas 462 empresas con representatividad provincial. Entre las variables consideradas, un grupo se corresponde con habilidades o capacidades para el uso de las TICs, las cuales fueron construidas como indicadores cuantitativos. En este caso, se empleó la técnica de Análisis de Componentes Principales y se identificaron dos dimensiones de relevancia: la primera caracteriza las habilidades derivadas de la experiencia de los trabajadores en la propia empresa relacionada al uso de PC e internet en sus tareas, al igual que el nivel de disponibilidad de PC; la segunda dimensión, por su parte, está determinada por la formación profesional individual de los trabajadores. Por otro lado, se tomaron en cuenta doce variables vinculadas a la infraestructura en TICs y los usos de Internet dentro de las empresas. Dada la naturaleza cualitativa de estas variables, se utilizó Análisis de Correspondencias Múltiples, definiéndose cuatro dimensiones de relevancia. La primera dimensión se constituye como un factor global de tamaño dada la alta correlación entre las variables, donde los usos de Internet como búsqueda de información y banca electrónica, junto a la disponibilidad de Intranet o Red de Área local, son las de mayor peso. La segunda dimensión engloba la conectividad a Internet y la comunicación online; la tercera, responde a aspectos comerciales: venta, compra y atención al cliente. Finalmente, la cuarta dimensión refiere principalmente al tipo de conexión a Internet. Esta primera caracterización de las dimensiones que constituyen la adopción de TICs por parte de las empresas, es una primera aproximación para la construcción de indicadores sintéticos que capten, entre otros aspectos, la complejidad intrínseca que significa la consideración de la infraestructura requerida, las políticas internas de las empresas y las habilidades y la formación necesarias por parte de los trabajadores.

Palabras claves: tics; componentes principa; análisis de correspo; empresas manufacture

INGRESANTES DE LA UNIVERSIDAD NACIONAL DE ROSARIO: PERFIL SOCIO-ECONÓMICO Y RENDIMIENTO ACADÉMICO

DIEGO MARFETÁN MOLINA, MARÍA EUGENIA TESSER, BEATRIZ LUCÍA MEINARDI y NERINA LÓPEZ

DIRECCIÓN GENERAL DE ESTADÍSTICA UNIVERSITARIA - UNR.

estadistica@unr.edu.ar

El estudio y la correcta interpretación de los perfiles sociales de los alumnos es una herramienta indispensable a la hora de lograr una eficiente gestión académica a nivel universitario. En este sentido, conocer las características de los propios estudiantes se convierte en un aspecto fundamental del diseño de cualquier plan estratégico, influenciando la toma de decisiones y definiendo las políticas a seguir por parte de las autoridades universitarias. Si bien este tema ha sido tratado con anterioridad en la Universidad Nacional de Rosario (UNR), principalmente en Moscoloni et al. (2007), el contexto social general y las realidades de los estudiantes de la UNR poseen una naturaleza dinámica que vuelve inevitable la permanente actualización de estos perfiles. En consecuencia, el principal objetivo de este trabajo es identificar grupos (clusters) de estudiantes con características similares, y al mismo tiempo analizar si estos atributos se encuentran correlacionados con el posterior rendimiento académico del alumno. En el presente trabajo se tiene en cuenta la cohorte de estudiantes ingresantes en el año académico 2014 a carreras de grado y pregrado dictadas en la UNR. Las variables analizadas están referidas a datos personales y al rendimiento académico de

los alumnos, obtenidas a partir de la información más actualizada y confiable a disposición de la Dirección General de Estadística Universitaria de la UNR. En un primer paso se transformaron las variables cualitativas a indicadoras, a partir de las cuales se construye una matriz de distancias entre alumnos utilizando el coeficiente de similitud definido por Gower (1971). Posteriormente se aplicó una técnica conocida como análisis de cluster jerárquico, la cual identifica grupos de estudiantes con características similares y los agrupa progresivamente en diversos clusters. Para llevar a cabo estos análisis se utiliza la librería MASS (Venables & Ripley, 2002) disponible en el paquete estadístico R. Los resultados del presente trabajo indican que, en líneas generales, no existe un perfil socioeconómico claramente definido que pueda ser asociado al rendimiento académico del alumno. No obstante, existen tendencias que sugieren una relación entre la deserción universitaria, la edad del ingresante y el área de estudio dentro de la cual se encuadra la carrera elegida. Además, variables referidas a la categoría ocupacional y al máximo nivel de estudios alcanzado por los padres también ejercen cierta influencia sobre el rendimiento académico de los alumnos de la UNR.

Palabras claves: cluster; rendimiento académico; perfil; ingresantes; unr

MODELO ESTRUCTURAL DE LAS RELACIONES ENTRE VARIABLES SOCIOECONÓMICAS Y LA DECISIÓN DE COMPRA FAMILIAR DE ALIMENTOS EN ADULTOS MAYORES

ANTONIO HUMBERTO CLOSAS¹ y JORGE GUILLERMO ODRIozola²

¹UNIVERSIDAD NACIONAL DEL NORDESTE - UNIVERSIDAD TECNOLÓGICA NACIONAL. ²UNIVERSIDAD NACIONAL DEL NORDESTE.

hclosas@hotmail.com

En un contexto globalizado, de economía de libre mercado, el consumo de productos en general, y la adquisición de alimentos para el ámbito familiar en particular, se presenta como un fenómeno multidimensional, modulado por diversos factores que van desde la disponibilidad económica hasta la influencia de los soportes comunicacionales. Estos determinantes participan a través de diferentes relaciones que inciden en forma directa e indirecta en la motivación, planificación y realización final de compra de las personas. A su vez, algunas características individuales de los consumidores, p. ej., los distintos segmentos sociales, ameritan que el estudio de esta temática se realice de manera diferenciada. Así pues, el objetivo del presente trabajo consiste en desarrollar mediante ecuaciones estructurales un modelo que explique de qué manera ciertos factores socioeconómicos influyen en la decisión de compra familiar de alimentos por parte de adultos mayores en ciudades del nordeste de Argentina. La investigación responde a un diseño explicativo, de estilo descriptivo mediante encuesta, de línea cuantitativa, de corte transversal y perfil correlacional. La muestra estuvo conformada por 460 individuos (263 mujeres, 197 hombres) de 60 a 85 años, que residen en las ciudades de Resistencia y Corrientes. En el procedimiento utilizado para extraer la muestra hemos combinado los métodos estratificado (las ciudades representan los estratos) y no probabilístico (la elección de las personas se realizó de manera intencionada e incidental). Las variables observadas o indicadores se han obtenido de aplicar un cuestionario socioeconómico, elaborado por un equipo interdisciplinario de académicos, investigadores y profesionales, con vasta experiencia en la región en relevamiento de consumo de diversos bienes y categorías de productos y servicios. Se propone un modelo conceptual que es contrastado a nivel empírico mediante la técnica Ecuaciones Estructurales, dando lugar a una representación ajustada a los datos, proveniente del procedimiento de estimación ERLS (Elliptical Reweighted Least Square), utilizando EQS, en la que intervienen seis constructos: Nivel socioeconómico, Situación en el hogar (independientes), Tecnologías de la información y la comunicación, Tiempo libre, Gastos (intermediarios) y Compras (dependiente). Tanto la metodología empleada, como el modelo final planteado, podrían ser recursos válidos para abordar con eficiencia la obtención de resultados y conclusiones que posibiliten comprender de manera integral, unificada

y cercana a la realidad una cuestión de central interés en sectores productivos, administrativos, institucionales, comerciales y empresariales del área objeto de estudio.

Palabras claves: modelización estadística; ecuaciones estructurales; consumo; adultos mayores; marketing

REDES NEURALES APLICADAS AL ESTUDIO DE LA CRIMINALIDAD EN EL DEPARTAMENTO DE CUNDINAMARCA, COLOMBIA

JULIAN DAVID REALPE CAMPAÑA y FREDSON GUERRERO

SERVICIO NACIONAL DE APRENDIZAJE - SENA.

jdrealpec@gmail.com

El crimen es un problema generalizado en todas las ciudades de Colombia y del mundo en general. Las causas y efectos del crimen en la sociedad han sido ampliamente estudiados y documentados, no obstante, en los últimos años con la ayuda de herramientas estadísticas computacionales, se han realizado importantes estudios que permiten brindar a las autoridades locales y sociedad en general de elementos alternativos de caracterización. En este trabajo se realizó el análisis estadístico de los datos de criminalidad entre los años 2008 a 2014 en el departamento de Cundinamarca, Colombia, mediante la utilización de índices de correlación, modelos lineales multivariados y redes neuronales. Los datos trabajados fueron obtenidos de parte de los boletines anuales del Observatorio de Seguridad del Departamento de Cundinamarca, Colombia. El cálculo del índice de correlación de los datos de criminalidad de las provincias mostró que los datos estaban espacialmente correlacionados, mostrando valores entre 0.85 y 0.99 para las provincias que compartían límites geográficos. Esto fue una indicación para utilizar modelos lineales multivariados que pudieran relacionar la ocurrencia de diversos tipos de delitos en cada una de las provincias de la Región de Cundinamarca con su cercanía geográfica. Los modelos lineales generados mostraron un ajuste promedio de 85% escogiendo para cada provincia los datos de criminalidad de sus vecinos inmediatos como variables independientes. Posteriormente, a fin de mejorar la capacidad predictiva, se utilizaron redes neuronales, las cuales se entrenaron con el 70% de los datos y probado con el 30% restante. La red neuronal mostró un valor de R-cuadrado de ajuste con los datos de prueba 0.95, mostrando una alta capacidad de predicción. Estos resultados apuntan a desarrollar metodologías alternativas para el estudio del crimen en la región, permitiendo a las autoridades locales proveer herramientas estadísticas que puedan ayudar a monitorear y predecir la distribución y evolución de la delincuencia en las distintas provincias de Cundinamarca.

Palabras claves: redes neuronales; multivariado; criminalidad; correlación espacial; cundinamarca; colombia

REDUCCIÓN DE DIMENSIÓN EN UN PROBLEMA DE HUMEDALES DEL PARANÁ MEDIO

STELLA MARIS VAIRA¹, LAURA MODINI², MARIEL ZERBATTO² y LETICIA MESA³

¹DEPARTAMENTO DE MATEMÁTICA - FBCB - UNL. ²SECCIÓN AGUAS - DPTO. DE CS. BIOLÓGICAS - FBCB - UNL. ³INSTITUTO NACIONAL DE LIMNOLOGÍA (CONICET-UNL).

stella.vaira@gmail.com

La ganadería extensiva, con modalidad rotacional, en humedales de la cuenca media del río Paraná,

es una estrategia de pastoreo común en las últimas décadas. Sin embargo, su efecto sobre la calidad bacteriológica de este sistema acuático es escasamente conocido. El objetivo de este trabajo fue analizar conjuntamente variables ($p = 6$) profundidad de la laguna, nivel hidrométrico del río, pH, cobertura de macrófitas, turbiedad y temperatura del agua en muestras ($n = 51$) de tres humedales que fueron considerados como factor, así como los indicadores bacteriológicos (Coliformes totales, Coliformes termotolerantes y *E. coli*) y la presencia o no de ganado, para la ubicación en el biplot de las nuevas componentes principales. La finalidad es identificar el número y composición de componentes necesarios para resumir las puntuaciones observadas en un conjunto grande de variables observadas; se aplicó un Análisis de Componentes Principales (ACP). Este método explica el máximo porcentaje de varianza observada en cada ítem a partir de un número menor de componentes que resume la información inicial. Los resultados muestran que la matriz de correlaciones iniciales tiene asociado un valor de covariancia generalizada de 0,095; el valor del estadístico KMO (Kaiser-Meyer-Olkin) de 0,382, si bien es bajo este valor, la prueba de esfericidad de Bartlett resultó altamente significativa. Los autovalores asociados a las dos primeras componentes dieron: 2,22 y 1,84; con un porcentaje acumulado de variancia del 68%. El biplot o representación punto-vector nos proporcionó la información de las variancias y covariancias de las variables y las distancias entre las muestras de agua. El ACP permitió distinguir, según los coeficientes de la combinación lineal, que las variables Temperatura, Nivel Hidrométrico y Turbiedad (propiedades del agua) explican la primera componente principal mientras la cobertura de macrófitas es la variable que mejor define la segunda componente principal. El software utilizado fue R.

Palabras claves: análisis de componentes; biplot; humedal; coliformes

REGIONES E INDICES SINTÉTICOS POR APLICACIÓN DE MÉTODOS ESTADÍSTICOS MULTIVARIADOS COMO BASE PARA LA COMPARACIÓN DE COMPONENTES DEMOGRÁFICAS

*MARTÍN SAINO, MARÍA CAROLINA TROGLIERO, DIEGO CONTRERAS y
JOSÉ CASTILLO SOSA*

INSTITUTO DE ESTADÍSTICA Y DEMOGRAFÍA, FACULTAD DE CIENCIAS ECONÓMICAS, UNIVERSIDAD NACIONAL DE CÓRDOBA.

martinsaino@gmail.com

Las condiciones de vida de una población están fuertemente ligadas tanto a la distribución espacial de las actividades económicas como a sus niveles de desarrollo regional, motivos por los cuales éstas también presentan diferenciales sociodemográficas. Conforme a estas perspectivas tales asimetrías pueden ser exploradas y/o explicadas merced a una adecuada selección de variables interactuantes, susceptibles al mismo tiempo de un correcto agrupamiento conforme niveles de homogeneidad. Así, sobre la base de un conjunto de indicadores elaborados con datos censales de 2001, y mediante la aplicación de Análisis de Componentes Principales (ACP) y Cluster se efectúa una regionalización y se construyen índices socioeconómicos sintéticos que reflejan la posición relativa de las provincias y esas regiones. Esos índices fueron construidos para cada jurisdicción a partir de la agregación de las componentes retenidas ponderadas por su peso y normalizadas por su máximo. Luego para cada región se tomó el promedio de los índices de las jurisdicciones que las conforman. Las regiones así constituidas y los índices de referencia sirven de base para la comparación de componentes demográficas en entornos básicamente desiguales. En ese sentido y en el marco de la transición demográfica, con información procedente de Estadísticas Vitales, se analizó niveles de mortalidad y el cambio en los patrones de fecundidad entre 2001 y 2015. A partir de los resultados obtenidos no cabe dudas que si bien globalmente Argentina se encuadra dentro de los países que presentan una etapa de transición avanzada, las regiones empero se encuentran todavía en diversos estadios de transición. Dan cuenta de ello las bajas tasas de natalidad como de mortalidad. En apoyo de estas

ideas vale la pena consignar que mientras la Región 1 está comenzando la Segunda Transición, las restantes regiones se ubican en distintos momentos y en estado avanzado de la tercera etapa de la Primera Transición.

Palabras claves: componentes principa; indices sintéticos; condiciones socioeco; componentes demográfaf

VARIABILIDAD EN LA EXPRESIÓN DE POLIPÉPTIDOS Y ATRIBUTOS FENOTÍPICOS DURANTE LA MADUREZ DEL TOMATE

A P DEL MEDICO¹, M S VITELLESCHI², A LAVALLE³ y G R PRATTA¹

¹IICAR (INSTITUTO DE INVESTIGACIONES EN CIENCIAS AGRARIAS DE ROSARIO), CONICET/UNR. ²IITAE (INSTITUTO DE INVESTIGACIONES TEÓRICAS Y APLICADAS DE LA ESCUELA DE ESTADÍSTICA), CIUNR/UNR. ³DEPARTAMENTO DE ESTADÍSTICA, UNIVERSIDAD NACIONAL DEL COMAHUE.

delmedico@iicar-conicet.gob.ar

Durante la madurez del fruto de tomate (*Solanum lycopersicum* L.) se producen cambios fisiológicos genéticamente regulados, que determinan un producto cuya calidad se adecue a los requerimientos de los consumidores. En esta transición se reconocen varios estados, de los cuales los más contrastantes son el Verde Maduro (VM, definido como aquél en el que el fruto inmaduro alcanza el tamaño máximo) y el Rojo Maduro (RM, en el cual el fruto presenta más del 90% de su superficie con el color de madurez). Una forma de evaluar la expresión génica diferencial es mediante el análisis de la Variabilidad de la Frecuencia con la cual determinados Polipéptidos (VFP) están presentes a lo largo de la madurez. Esta variabilidad en el nivel molecular podría estar relacionada con la Variabilidad para Atributos Fenotípicos (VAF) del fruto. El objetivo de este trabajo fue estudiar, mediante Análisis Factorial Múltiple (AFM), la posible asociación de VFP en VM y RM con VAF. El material vegetal estuvo compuesto por 75 individuos de tomate de la generación segregante F2 del híbrido de segundo ciclo F1 ToUNR15xToUNR9. De cada individuo se cosecharon 2 frutos en VM y 2 frutos en RM, de cuyo pericarpio se extrajeron los polipéptidos totales. Los perfiles moleculares se determinaron por electroforesis convencional. Para evaluar VFP, se seleccionaron en VM y RM polipéptidos con frecuencias extremas (0 y 1) y que ajustaron a una segregación mendeliana conocida. Para evaluar VAF, en cada individuo se midieron peso, diámetro, altura, forma y vida poscosecha. Se aplicaron dos AFM, el primero para relacionar VFP en VM con VAF y el segundo para relacionar VFP en RM con VAF. Los resultados de ambos AFM se compararon a fin de verificar si alguno de los estados de madurez presenta mayor asociación con los atributos fenotípicos. En VM, 12 polipéptidos cumplieron los requisitos para ser seleccionados, mientras que 7 polipéptidos lo hicieron en RM. La proporción de variación explicada por los dos primeros ejes principales fue mayor en el AFM correspondiente a RM (39,8% vs. 28,1% en VM). Las representaciones de los individuos y de las variables en función de VFP y VAF difirieron entre VM y RM, lo que indica que la asociación entre la variabilidad molecular y la fenotípica fue diferente en ambos estados. Sin embargo, el coeficiente RV fue 0,05 en VM y 0,06 en RM, evidenciando una baja magnitud de tales asociaciones.

Palabras claves: análisis factorial m; *solanum lycopersicum*; expresión génica

Métodos no paramétricos

COMPARACIÓN DEL ORDENAMIENTO DE PREDICCIONES BLUP EN MODELOS DE EVALUACIÓN GENÉTICA PARA CONEJOS

E N FERNÁNDEZ¹, N N ABBIATI¹, J P SÁNCHEZ², R D MARTÍNEZ¹, C P
GONZALEZ¹ y M S ROVEGNO¹

¹FACULTAD DE CIENCIAS AGRARIAS, IIPAAS, UNIVERSIDAD NACIONAL DE LOMAS DE ZAMORA. ²GENETICA I MILLORA GENETICA ANIMAL, INSTITUT DE RECERCA I TECNOLOGIA AGROALIMENTÀRIES.

ednfer@yahoo.com

El modelo mixto aditivo con repetibilidad es el más empleado para mejorar el número de gazapos destetados (ND) en conejos para carne. Estudios recientes detectaron una fuerte asociación entre el efecto fijo año-estación (AE) y la consanguinidad, que condujo a una sobreestimación de parámetros y tendencias genéticas. El AE incluido como aleatorio en el modelo produjo estimaciones más razonables. El objetivo de este trabajo fue comparar el ranking de los valores genéticos aditivos predichos según se considere el AE fijo o aleatorio en el análisis del carácter ND en conejas de la línea A pertenecientes a la Universidad Politécnica de Valencia. Se dispuso de una base de 15671 registros para ND y de una genealogía de 5668 conejos que comprendieron 38 generaciones. Los datos fueron analizados por dos modelos, diferenciados por la inclusión de AE como fijo (M1) o aleatorio (M2). Además, en ambos se consideró como efecto fijo al estado fisiológico de la hembra y la covariable consanguinidad, y como efectos aleatorios el valor genético aditivo y el efecto permanente de la coneja. La relación entre los rankings de ambos modelos se estudió empleando dos aproximaciones: la correlación de Spearman (para todas y las veinte o cuarenta mejores conejas (T20 o T40), en cada generación) y el porcentaje de discordancias en la elección de las T20/T40, según generación. Al considerar la generación completa, las correlaciones fueron altas con un promedio por generación de 0.99 (± 0.01), disminuyendo a 0.90 (± 0.08) y 0.94 (± 0.03) en las T20 y T40 respectivamente. El porcentaje de discordancias promedio en T20 fue 8.55% ($\pm 5.05\%$) y de 6.18 % ($\pm 3.3\%$) en T40. Se detectaron generaciones con el mismo porcentaje de discordancia y correlaciones de Spearman diferentes, como en las generaciones 5 y 11 de T20 con un porcentaje de discordancia del 5% y correlaciones de 0.72 y 0.99, respectivamente. También se presentaron generaciones con % de discordancia similares y correlaciones diferentes. Sin embargo, la respuesta estimada a la selección a través de las hembras elegidas por los M1 o M2, evaluadas por M2, fue de 0.05 gazapos por generación tanto para las T20 como las T40. En nuestro caso y poniendo énfasis en los resultados del proceso selectivo, el estudio de los rankings entre modelos mediante correlaciones y % de discordancias no resultó concluyente respecto a la respuesta esperada a la selección.

Palabras claves: modelos mixtos; coeficiente de spear; respuesta a la selec; cunicultura

Métodos robustos

ANÁLISIS DE VARIANZA ROBUSTO EN ENSAYOS DE NUTRICIÓN EN POLLOS PARRILLEROS

DIANA ALEJANDRA GIORGINI, STELLA MARIS ZABALA, SUSANA FILIPPINI y LUCIANO PALACIOS

dgiorgin@agro.uba.ar

La producción avícola en la República Argentina afronta actualmente la prohibición del uso de antimicrobianos como promotores del crecimiento con el empleo de acidificantes, enzimas, antioxidantes, probióticos, beta-adrenérgicos, pigmentadores, estimulantes y extractos vegetales, con el propósito de mantener en niveles aceptables la productividad y eficiencia. El objetivo de esta investigación es estimar y predecir diferencias en el comportamiento de un parámetro productivo en el área avícola, a partir del empleo del Modelo Lineal General (ML) para un análisis de Varianza, en ensayos que pongan a prueba enzimas fitasas. El ensayo se realiza en la Granja Experimental de una empresa ubicada en San Andrés de Giles, utilizándose pollos parrilleros en condiciones de crianza comercial. Se aplicó un diseño Completamente Aleatorizado. La aleatorización, fue realizada considerando, 16 (dieciséis) repeticiones con 4 (cuatro) tratamientos. Se comprobó la homogeneidad de Variancias entre tratamientos. Asimismo, se analizó por métodos robustos si existían datos outliers. Cuando se encontraron, se aplicaron Métodos Robustos o Técnicas de acomodación de los datos anómalos. Estos tienen mayor sensibilidad ante la presencia de datos anómalos o outliers. Están diseñados para realizar inferencias sobre el modelo, reduciendo la posible influencia que pudiera tener la presencia de datos anómalos. La información obtenida se estudió con un Análisis de Variancia para un diseño en Completamente aleatorizado y por métodos robustos. Se utilizaron además de los Métodos Clásicos como la prueba de Tukey ó Dunnett (existe un control o testigo), pruebas robustas para las comparaciones entre grupos y dentro de grupo cuando no existía homogeneidad de variancias, analizando la concordancia entre todas las pruebas. Se deja constancia que las pruebas se plantearon con un nivel de significación del cinco por ciento.

Palabras claves: datos outliers; análisis de varianza; métodos robustos

UN ESTUDIO SOBRE INDICADORES DE CARENCIAS CRÍTICAS A TRAVÉS DE UN ANÁLISIS DE COMPONENTES PRINCIPALES ROBUSTAS

PATRICIA CICCIOLI y JAVIER BUSSI

INSTITUTO DE INVESTIGACIONES TEÓRICAS Y APLICADAS DE LA ESCUELA DE ESTADÍSTICA, UNIVERSIDAD NACIONAL DE ROSARIO.

ciccioli.patricia@gmail.com

Los indicadores socio-económicos de carencias críticas para ciudades y comunas de la provincia de Santa Fe permiten caracterizar estas poblaciones de acuerdo a 10 variables que describen aspectos relacionados con la conformación y las condiciones presentes en los hogares, como así también características de las viviendas y cuestiones referidas a escolaridad y situación laboral de la población joven. Estas variables se refieren a 158 ciudades y comunas de más de 2000 habitantes de la provincia de Santa Fe, los cuales provienen del Censo Nacional de Población, Hogares y Viviendas

2010. Para poder describir y resumir las diferencias existentes entre las ciudades y comunas con respecto a estos indicadores socio-económicos es posible utilizar el Análisis de Componentes Principales (ACP) que permite representar un conjunto de n observaciones con p variables a través de un número menor de variables construidas como combinaciones lineales de las originales y conservar la mayor variabilidad posible de los datos. Este método clásico utiliza medidas de variabilidad (como por ejemplo, la matriz de variancias y covariancias), las cuales están influenciadas por la presencia de valores extremos (outliers), lo que puede generar una distorsión en la representación. Debido a esta cuestión, en este trabajo se presentan dos métodos robustos para el ACP, el correspondiente a Matriz de Covariancia de Determinante Mínimo (Minimum Covariance Determinant, MCD), y el método de Componentes Principales Esféricas (Spherical Principal Components, SPC). El objetivo de este trabajo es comparar estos dos métodos con el método clásico aplicándolos a estos indicadores socio-económicos. Para poder resumir las diferencias existentes entre las ciudades y comunas es necesario retener un número mayor de componentes principales en los métodos robustos (MCD y SPC) que en el método clásico. Se nota además que las variables que más influyen para resumir las diferencias de carencia social-económica son las mismas, siendo en total 4 de las 10 originales: Porcentaje de jefes de hogar con educación primaria incompleta, Porcentaje de hogares sin caño de agua dentro de la vivienda, porcentaje de hogares en vivienda sin retrete con descarga de agua y Porcentaje de población de 14 a 19 años que asiste a nivel de instrucción primario. Sin embargo, estas variables tienen una distinta participación en las componentes principales, según el método utilizado.

Palabras claves: indicadores carencia; componentes principa; minimum covariance d; spherical principal

Minería de datos

IDENTIFICACIÓN DE PERFILES DE LA VOZ EN ENFERMOS DE PARKINSON CON REDES NEURONALES

*GASTÓN BERRETTA¹, MONICA GIULIANO¹, EVANGELINA MALDONADO²
y GABRIEL BLANCO¹*

¹DEPARTAMENTO DE INGENIERÍA E INVESTIGACIONES TECNOLÓGICAS. UNIVERSIDAD NACIONAL DE LA MATANZA. ²DEPARTAMENTO DE CIENCIAS DE LA SALUD. UNIVERSIDAD NACIONAL DE LA MATANZA.

mgiuliano@unlam.edu.ar

Vinculación entre parámetros de la voz y avance de la enfermedad en pacientes con enfermedad de Parkinson (EP). Si bien el diagnóstico definitivo es histopatológico, se considera que un diagnóstico preliminar en etapas tempranas resulta ser un desafío y una enorme responsabilidad, considerando las implicancias de pronóstico vital y las limitaciones motrices futuras para cada paciente diagnosticado. En esta dirección, se orientan múltiples investigaciones que hacen uso de Tecnologías de la Información y las Comunicaciones en el área de salud, en este caso se trata de predecir el grado de avance de la EP a través de parámetros de la voz grabada de pacientes. A partir de una base pública de enfermos ya diagnosticados (747 observaciones) donde se ha grabado un fonema y se han parametrizado matemática a través de 38 parámetros, de los cuales seleccionamos 36. A su vez el grado de enfermedad está incluido en la base según las respuestas a una versión reducida de la escala PDERS (ParkinsonsDisease Rating Scale) y mediante el valor suma de esta escala, que dividimos en 2 categorías. En nuestro trabajo realizamos 3 tipos de análisis con Redes Neuronales. Se plantea un primer acercamiento al problema con una red neuronal perceptrón multicapa con la

siguiente arquitectura: capa de entrada, capa de salida y dos capas ocultas Optimizando y revisando los parámetros, llegamos a una configuración que, utilizando la base de datos completa resultó en una clasificación correcta del 89,96%. Luego, manteniendo la configuración, decidimos armar el modelo con un 50% de los datos y evaluando el modelo con el restante 50%, con lo que obtuvimos un 64,34% bien clasificado. Por último armamos el mismo modelo en base a una validación cruzada de 10 iteraciones, obteniendo un 61.32% correctamente clasificado. En base a estos resultados, realizaremos un análisis exploratorio para profundizar en los casos que fueron predichos correctamente, analizando las variables involucradas y descubriendo si existen patrones de comportamiento en estas variables.

Palabras claves: enfermedad de parkin; voz; pdrs; redes neuronales

PONDERACIÓN DE LA TÉCNICA “SUPPORT VECTOR MACHINE” (SVM) PARA EL MEJORAMIENTO EN LA PREDICCIÓN DE LA DISTRIBUCIÓN DE TAMAÑOS A COSECHA DE FRUTOS DE PERAS CULTIVAR BEURRE D’ANJOU

GUSTAVO GIMÉNEZ¹, NATALIA RUBIO¹, DOLORES DEL BRÍO² y SERGIO BRAMARDI¹

¹DEPARTAMENTO DE ESTADÍSTICA, FACULTAD DE ECONOMÍA, UNIVERSIDAD NACIONAL DEL COMAHUE. ²CONSEJO NACIONAL DE INVESTIGACIONES CIENTÍFICAS Y TÉCNICAS-INSTITUTO NACIONAL DE TECNOLOGÍA AGROPECUARIA..

gustavo.gimenez@fAEA.uncoma.edu.ar

La técnica denominada “Support Vector Machine (SVM)” se basa en la teoría de “statistical learning” y consiste en aplicar un método lineal sobre un conjunto de datos en un espacio de alta dimensionalidad. Es una técnica que se aplica a distintos problemas del aprendizaje estadístico como: clasificación, regresión y detección de datos nuevos (“Novelty Detection”). La SVM se ha aplicado en pronóstico de diversos cultivos con resultados muy precisos. Su aplicación no requiere conocer la relación matemática entre variables de entrada (predictoras) y de salida (respuesta) ni la distribución de probabilidad subyacente y es aplicable a enormes bases de datos, con bajo costo computacional. El presente trabajo retoma un conjunto de curvas de crecimiento de peras D’Anjou medidas en milímetros y evaluados en días a partir de la floración donde cada fruto fue categorizado por el tamaño comercial a cosecha. En dichos datos se implementó SVM 30 y 40 días previos a la cosecha para predecir los tamaños de los frutos a partir del diámetro de los mismos y su clasificación en tamaños comerciales. El SVM se calibró con un kernel lineal y el parámetro $C = 1$. El parámetro C controla la penalidad por las clasificaciones erróneas del SVM en datos de entrenamiento. Cuanto mayor es el valor de C , mayor la reducción de los errores de clasificación, incrementando la rigidez del algoritmo y la posibilidad de sobreajustar los datos. De esta forma se logró una precisión (“accuracy”) de 0.60, una tasa de no información de 0.26 y el test de no información vs precisión un p-valor <0,0001. No obstante, la matriz de confusión mostró que la predicción de frutos grandes y pequeños alcanzaba valores de 0.54 y 0.5 respectivamente. Una solución a este desbalance fue ponderar los vectores soporte por las proporciones de los tamaños comerciales de los datos de entrenamiento. Las ponderaciones permiten sesgar el modelo afectando el parámetro C para cada categoría de clasificación y compensar las categorías que están menos representadas en los datos. Éstas pueden ser realizadas según distintos criterios: $1/w$, $w-1/2$, $1500*w-1/2$, donde w es la proporción de cada categoría. En los datos de peras cultivar D’Anjou, el criterio de ponderación que obtuvo mayor precisión fue $w -1/2$ alcanzando el 0.63, mejorando la precisión en los tamaños pequeños y grandes a valores de 0.93 y 0.81 respectivamente, muy superiores a los valores sin ponderar.

Palabras claves: datos multicategóricos; “statistical learning”; penalización; vectores soporte

SELECCIÓN DE ATRIBUTOS MEDIANTE ENTROPÍA CONDICIONAL PARA DIFERENTES NIVELES DE INCERTIDUMBRE

MARÍA DEL CARMEN ROMERO¹ y ALEJANDRO CLAUSSE²

¹FACULTAD DE CIENCIAS ECONÓMICAS, UNIVERSIDAD NACIONAL DEL CENTRO DE LA PROVINCIA DE BUENOS AIRES. ²PLADEMA, FACULTAD DE CIENCIAS EXACTAS, UNIVERSIDAD NACIONAL DEL CENTRO DE LA PROVINCIA DE BUENOS AIRES.

mariadelc.romero@gmail.com

La clasificación supervisada y la selección de atributos diferenciales tienen múltiples aplicaciones; en particular en contextos de alta dimensionalidad, debido al crecimiento de las bases de datos (no sólo por la cantidad de observaciones sino también de atributos, y por los escenarios con mayor cantidad de atributos que de observaciones). En Romero, Di Rienzo y Clausse (2013) se presentó un método que usa la entropía condicional para seleccionar atributos que pueden considerarse diferenciales en la determinación de grupos (dados por un atributo de clasificación) y en Romero y Clausse (2016) se analizó el método para diferente cantidad de observaciones y de atributos, considerando distintos niveles de dependencia. La entropía condicional se define como la cantidad media de incertidumbre que tiene un atributo (en este caso, el que determina el grupo) conociendo la información que aporta otro. Si la entropía del grupo disminuye por el conocimiento de algún atributo, éste es relevante para la distinción entre grupos. Una mejora al método propuesto consiste en recomendar un determinado subconjunto de atributos diferenciales considerando la máxima incertidumbre que se esté dispuesto a tolerar (en términos de entropía condicional). En general, es esperable que ante altos niveles de incertidumbre, unos pocos atributos sean suficientes para distinguir entre grupos, mientras que ante menores niveles, se requiera una cantidad mayor. En este trabajo, se aborda la problemática de detectar estos diferentes puntos de corte, en los cuales se modifica la cantidad de atributos diferenciales a considerar. El objetivo es explorar preliminarmente distintos indicadores presentes en los conjuntos de datos que puedan ser utilizados para “detectar” estos puntos de corte. Entre los principales resultados, se encuentra que la entropía de la cantidad de atributos seleccionados es un indicador pertinente y que presenta un buen comportamiento incluso en contextos con una baja cantidad de observaciones. Este indicador se calcula realizando permutaciones sobre el conjunto de datos original. Se trabajó con escenarios de atributos cualitativos, en particular binarios, en los cuales la selección de diferente cantidad de atributos diferenciales conduce a distintos niveles de incertidumbre en la determinación del grupo. Se utilizó el software R y los diferentes escenarios fueron generados mediante simulación.

Palabras claves: selección de atribut; clasificación superv; entropía; simulación

Modelos de regresión

ANÁLISIS DE ESTRÉS EN POTRILLOS AL APLICAR LAS TÉCNICAS DE VINCULACIÓN Y APRENDIZAJE NO TRAUMÁTICO EMPLEANDO MODELOS MIXTOS

N ABBIATI¹, S PAZ², M TOPAYAN², D REFOJO² y O MEDINA¹

¹CÁTEDRA DE BIOMETRÍA, FACULTAD DE CIENCIAS AGRARIAS, UNLZ. ²CÁTEDRA DE EQUINOTECNIA, FACULTAD DE CIENCIAS AGRARIAS, UNLZ.

norabbi2000@gmail.com

Las Técnicas de Vinculación y Aprendizaje No Traumático (TVANT) son un conjunto de maniobras realizadas a potrillos entre los días 7 y 14 de nacidos, para establecer un vínculo hombre-animal; generando desensibilización a situaciones estresantes, que impiden una sociabilización temprana. Las variables frecuencia cardíaca y temperatura rectal, después de actividades específicas, permiten cuantificar el estrés. El objetivo del trabajo fue estudiar si existe un incremento en dichas variables luego de aplicar las TVANT en condiciones a campo. Se seleccionaron aleatoriamente 21 potrillos de raza criolla, de un haras de Luján, y 5 Silla Argentino, del campo experimental Santa Catalina de la UNLZ. Las taras se efectuaron en instalaciones aptadas para tal fin, muy diferentes, en condiciones de campo. Entre los días 7 a 14, antes y después del tratamiento, se midió temperatura rectal con termómetro digital y frecuencia cardíaca con estetoscopio Littman. El análisis de la información se realizó mediante Modelos Mixtos para contemplar una estructura de correlaciones entre mediciones de un mismo animal y/o heterogeneidad de varianzas. El modelo contempló los factores: Raza-Lugar (Criollos-Luján, CLu, Silla Argentino-Santa Catalina SASC); Momento (previo y posterior a las TVANT) y Día (7 a 14) con sus interacciones dobles y triple. Para la elección de la estructura de covarianzas de la matriz de errores se empleó el criterio de información de Akaike; para la comparación de las medias mínimo cuadráticas, Tukey-Kramer. Se usó el software SAS con $\alpha=0,05$. Con relación a temperatura, el modelo elegido contempló heterogeneidad de varianzas para Raza-Lugar y detectó diferencias entre momentos siendo el momento previo inferior al posterior. También detectó interacción Raza-Lugar y Día. Se efectuó la apertura de Día dentro de Raza-Lugar, detectándose en CLu una disminución significativa de medias entre los días 7 y 10; para SASC detectó una disminución entre la temperatura media del día final y los días iniciales (7, 8 y 10). En frecuencia cardíaca, el modelo contempló una matriz desestructurada para el error y detectó interacción Raza-Lugar y Día. Al efectuar la apertura de Día dentro de Raza-Lugar, en CLu no había diferencias significativas para los dos primeros, pero disminuían y se mantenían estables a partir del día 9. Para SASC, se presentaron diferencias entre la totalidad de días (excepto 9) y el último. Se concluye existencia de comportamiento diferencial Razas-Lugar, con disminución en las medias de ambas variables a lo largo del tiempo; presentándose mayor estabilidad en CLu.

Palabras claves: modelos mixtos; potrillos; técnicas de sociabil

ANÁLISIS DE LA EFICIENCIA DE DIFERENTES DOSIS DE ENZIMAS CON MODELOS LINEALES GENERALIZADOS EN ENSAYOS DE NUTRICIÓN EN POLLOS PARRILLEROS

LUCIANO PALACIOS, MARÍA LAURA VIGNERA, MAURICIO DE
FRANCESCHI y SUSANA FILIPPINI

UNLU.

lucianofepalacios@hotmail.com

Análisis de la eficiencia de diferentes dosis de enzimas con Modelos Lineales Generalizados en ensayos de nutrición en pollos parrilleros PALACIOS, L.1, VIGNERA, L.1., DE FRANCESCHI, M.2, FILIPPINI, S.1 1División Estadística, Departamento Cs. Básicas, Universidad Nacional de Lujan 2Avicultura, Departamento Tecnología, Universidad Nacional de Lujan lucianofepalacios@hotmail.com Resumen El empleo de modelos matemáticos para la explicación de fenómenos probabilísticos ha sido imprescindible en la investigación científica. No obstante, es frecuente tener que trabajar con variables que no cumplen con las características requeridas por el Modelo Lineal General, siendo el modelo Lineal generalizado, el que responde adecuadamente a los problemas generados por la métrica de las variables. El objetivo de este trabajo es modelar la eficiencia de diferentes dietas de nutrición en pollos parrilleros, a través del uso del indicador FEP (Factor de Eficiencia de Producción), cuyo cálculo es el siguiente: FEP es igual al producto entre el Peso y la Viabilidad de

grupo de animales, dividido por el producto entre la Conversión alimenticia y la Edad del lote, todo multiplicado x 100. El criterio para dicotomizar los resultados de este índice se llevó a cabo, estipulando como punto de corte el valor teórico de FEP=390 que debería obtener un grupo de pollos parrilleros a los 42 días de edad, 2.953 kg de peso y 1.69 de conversión alimenticia. Estos parámetros son prestablecidos por el proveedor de la línea genética Cobb a través del Suplemento Informativo Sobre Rendimiento Y Nutrición De Pollos De Engorde Cobb500. Se observa la falta de ajuste de los residuos a la media, es decir el no cumplimiento de la independencia de las observaciones. Se prueban diferentes modelos Lineales Generalizados de dosis respuesta y funciones de enlace como Probits, Logit y Log- log complementario. Se emplea el análisis de las Devianzas de los distintos modelos, como una generalización del Análisis de la Varianza para una secuencia de modelos, cada uno incluyendo más términos que los anteriores-Se comparan las bondades de ajustes con el Criterio de Información de Akaike. Se detectan diferencias significativas entre los tratamientos que poseen diferentes dosis de enzimas fitasas a partir de la comparación de medias con Bonferroni secuencial. Se concluyó que el mejor tratamiento es el que posee una dosis de enzima fitasa a razón de 2500 FTU con diferencias estadísticas sobre las dietas que no poseen.

Palabras claves: enzimas; modelos lineales gen; devianzas; bonferroni secuencia

ANÁLISIS DE LA INFORMACIÓN FINANCIERA Y DE MERCADO EN EMPRESAS LATINOAMERICANAS A TRAVÉS DE LA APLICACIÓN DE MODELOS MIXTOS

LETICIA TOLOSA¹ y MARÍA PAULA ROJO²

¹UNIVERSIDAD NACIONAL DE CÓRDOBA-FACULTAD DE CS.ES.. ²UNIVERSIDAD NACIONAL DE CÓRDOBA-FACULTAD DE CS.ES.

leticiaetolosa@gmail.com

Resumen Ejecutivo ¿Cuáles son los indicadores que explican los retornos de las acciones en el mercado de capitales? Según Chaopricha (2007) la respuesta a esta pregunta ha sido muy buscada por los investigadores e inversionistas, y la mayoría de los estudios hasta la fecha son de naturaleza empírica y proporcionan una visión muy limitada del problema. Las conclusiones generales son, que ciertas características de la empresa sí tienen un efecto en las existencias de retornos. Entre estas características, las más estudiadas incluyen relación de precio-valor libro, tamaño y precio-ganancia. En el presente trabajo se realiza, en una primera etapa, un estudio descriptivo del mercado latinoamericano a los fines de mostrar la situación relativa de cada país, con el fin de identificar las particularidades sobre cantidad de empresas cotizantes y capitalización bursátil de Argentina, Chile, Perú, Colombia y Brasil. A continuación en una segunda etapa, a través de la aplicación de modelos estadísticos, se analizan datos de mercado e información contable de empresas no financieras que conforman los índices bursátiles más representativos de los países mencionados anteriormente. El horizonte temporal de análisis es 2011 a 2016. Para explicar la variación de precios de las acciones en el período mencionado, el panel de datos conformado de cada país se analiza a través de la aplicación de modelos lineales mixtos (Tolosa, 2013). Los resultados obtenidos de acuerdo a los datos muestrales, evidencian que para explicar la variación de precios de las acciones en el mercado, resulta significativo y con relación directa la Rentabilidad del Patrimonio Neto (RPN) en el mercado de Brasil, Perú, y Chile, aunque el ratio de Solvencia (RSOL) solo resulta significativo en forma directa en Brasil. En el mismo sentido el cociente entre el Precio y el Valor de Libros (PVL), resultan significativo para los mercados de Argentina, Brasil y Chile. El tamaño de las empresas medido a través de la Capitalización Bursátil (CBU) es significativo de manera inversa en Argentina y Colombia. El Año resulta significativo para Argentina y Chile considerado como efecto fijo y en todos los modelos se lo incluye como efecto aleatorio. El ratio precio – ganancia (PE) no aparece como significativo en ninguno de los mercados analizados en este espacio temporal Al no resultar estadísticamente significativos indicadores comunes para las empresas domésticas de cada país, es

de interés el análisis de las diferencias observadas, con el resultado del análisis descriptivo realizado en la primera parte del trabajo.

Palabras claves: modelos mixtos; indicadores contable; indicadores de merca; retornos trimestrales

APLICACIÓN DE MODELOS LINEALES GENERALIZADOS MIXTOS PARA LA EVALUACIÓN DE EFFECTOS DE MEZCLAS BINARIAS DE PLAGUICIDAS

P M DEMETRIO

CENTRO DE INVESTIGACIONES DEL MEDIO AMBIENTE, FACULTAD DE CIENCIAS EXACTAS, UNLP. LA PLATA, BUENOS AIRES, ARGENTINA.
CONICET.

pablo.demetrio@gmail.com

La Ecotoxicología como disciplina tiene a los bioensayos de toxicidad como una de las herramientas fundamentales. Esta herramienta ha sido utilizada ampliamente para evaluar el efecto biológico de un tóxico y, posteriormente, mezclas de tóxicos sobre distintos organismos modelo. En la región pampeana, la ocurrencia simultánea de plaguicidas en cuerpos de agua superficiales es un escenario de exposición recurrente para la biota acuática; llegando estos xenobióticos a partir de la escorrentía superficial y/o deriva de las zonas de cultivos. Por esta razón la evaluación del impacto de mezclas de plaguicidas que se presentan con alta frecuencia en estos tipos de sistemas es ambientalmente relevante. En la actualidad, dentro de la Ecotoxicología se encuentran en discusión las metodologías implementadas para el análisis de datos de los bioensayos de toxicidad y de las mezclas de sustancias en particular. El objetivo del presente trabajo fue evaluar de manera preliminar la toxicidad de la mezcla de los herbicidas glifosato y 2,4-D sobre el cladócero *Daphnia magna* en ensayos estáticos a 48 h en condiciones controladas de laboratorio. Se realizaron pruebas de cada herbicida por separado y posteriormente con las mezclas binarias. El análisis de resultados de los bioensayos de los herbicidas individuales se realizó en el contexto de los Modelos Lineales Generalizados (MLG) y en el caso de la mezcla binaria en el contexto de los Modelos Lineales Generalizados Mixtos (MLGM). En estos últimos se evaluaron modelos con y sin interacción entre los herbicidas. Como ventajas de este abordaje en relación a los tradicionales utilizados cabe destacar que: a) no es necesaria la transformación de la variable binomial, b) es posible utilizar criterios estadísticos para la elección de los modelos (aditividad vs interacción) y c) es viable incorporar la variabilidad inter-ensayos, permitiendo modelarla debido a la no simultaneidad temporal de las pruebas. Los resultados indican que no hay evidencia de interacción de los herbicidas glifosato y 2,4-D para las condiciones experimentales evaluadas, considerando un efecto aditivo sobre la supervivencia de *D. magna*. El abordaje del efecto de las mezclas mediante GLM y sus extensiones es una alternativa para este tipo de estudios en el ámbito de la Ecotoxicología. Agradecimiento: PICT 2013-2393

Palabras claves: ecotoxicología; mezclas; glifosato; 2,4-d; mlg; mlgm

APLICACIÓN DE MODELOS LINEALES MIXTOS PARA EVALUAR LA CAPACIDAD PULMONAR DE ESCOLARES DE LA CIUDAD DE ARTIGAS (URUGUAY) Y LA RELACIÓN CON SU ENTORNO 2015-2016

RAMÓN ÁLVAREZ¹, YOHANA ALTEZ DE CASTRO² y VIRGINIA BURGUETE EGUREN²

¹ INSTITUTO DE ESTADÍSTICA - DMMCC - FCEA. ².

altezyohana@gmail.com

Los modelos lineales mixtos son una herramienta eficaz que permite el análisis de datos longitudinales. En la presente investigación se utilizará esta herramienta para analizar la capacidad pulmonar de los niños de la ciudad de Artigas (Uruguay), en cuatro etapas de monitoreo a lo largo de un año, contemplando períodos de intensificación de la actividad industrial (dentro o fuera de zafra) de la zona. **OBJETIVOS:** Evaluar diferentes tipos de modelos lineales mixtos para datos correlacionados temporal y espacialmente, que permitan evaluar si la capacidad pulmonar medida a través del pico flujo espiratorio se asocia con la edad, sexo, zona geográfica y la etapa de monitoreo, entre otras. **MÉTODOS:** El estudio consiste en monitorear la funcionalidad respiratoria de una muestra representativa de la población, que consta de 714 niños de 4 a 7 años inclusive, matriculados en las escuelas urbanas de la ciudad de Artigas. Se analiza la evolución de los individuos muestreados evaluando su capacidad pulmonar y su relación con otras variables. Se realizan comparaciones de distintos modelos lineales mixtos en los cuales se quiere explicar una variable cuantitativa (capacidad pulmonar) y su relación lineal con factores fijos, factores aleatorios y las covariables. **RESULTADOS PRELIMINARES:** Se analizaron perfiles medios y patrones de no respuesta, así como trayectorias individuales de cada escolar. **CONCLUSIONES:** Mediante el análisis descriptivo se muestra una diferenciación notoria entre niñas y varones sin identificar patrones con respecto a la zafra. Tampoco se aprecian asociaciones entre zona de residencia con respecto a la capacidad pulmonar, las cuales sí podrían detectarse con la aplicación de modelos lineales mixtos.

Palabras claves: capacidad pulmonar; datos longitudinales; modelos lineales mix

APLICACIÓN DEL MODELO DE REGRESIÓN LOGÍSTICA PARA EL ESTUDIO DE LA DESERCIÓN DE LOS INGRESANTES 2016 DE LAS CARRERAS DE GRADO DE LA FACULTAD DE CIENCIAS ECONÓMICAS DE LA UNER

OLGA BEATRIZ AVILA y STEFANÍA D'IORIO

FACULTAD DE CIENCIAS ECONÓMICAS- UNER.

olga.beatriz.avila@gmail.com

El problema de la deserción estudiantil es uno de los problemas más complejos y frecuentes que enfrentan las Universidades de nuestro país, Argentina, lo que resulta en la necesidad de realizar investigaciones para identificar los factores que influyen en el abandono de los ingresantes a las carreras universitarias. Los resultados de estas investigaciones servirían para identificar y atender las causas que intervienen en el abandono, y poder ejecutar políticas universitarias tendientes a la retención de los ingresantes. En este marco, el presente trabajo tuvo como objetivo identificar los factores que influyen en la deserción de los ingresantes 2016 a las carreras de grado de la Facultad

de Ciencias Económicas de la Universidad Nacional de Entre Ríos. Para ello se utilizó un modelo de regresión logística donde se considera la influencia de diferentes variables en la reinscripción de los estudiantes al segundo año de la carrera, en este caso, la reinscripción al año académico 2017, considerando esta variable como posible indicador de deserción. El modelo logístico resulta apropiado cuando la variable dependiente es dicotómica. También llamado modelo de respuesta cualitativa, tiene utilidad para pronosticar qué sucederá cuando existan dos posibilidades, en nuestro caso, se reinscribió o no al siguiente año académico en dos de las carreras que se dictan en la Facultad, que son Contador Público y Licenciatura en Economía. Como posibles variables predictoras se consideraron, entre otras, género, materias aprobadas, cantidad de materias cursadas en el año 2016 y edad al finalizar el periodo. Se trabajó con el total de ingresantes a las carreras mencionadas del año académico 2016, que fueron 375 estudiantes. Por un lado, para ver la de bondad de ajuste del modelo se consideró la prueba de Hosmer y Lemeshow, que indicó un buen ajuste del modelo. Por otro lado, del modelo obtenido cuyas variables estadísticamente significativas resultaron materias aprobadas, materias aprobadas el año anterior, cantidad de materias cursadas en el año 2016, edad al finalizar el periodo, se obtuvo una clasificación para los casos estudiados del 81,6%. Es necesario continuar con el estudio de las bases de datos de que se dispone, correspondientes a los alumnos ingresantes, para detectar un mayor número de variables que permitan identificar posibles causas de malos rendimientos en el área y, en consecuencia, tomar las medidas académicas necesarias para atenuar esta problemática. En este sentido, se debe mencionar que la Facultad cuenta con un Gabinete Psicopedagógico que trabaja en estos aspectos.

Palabras claves: regresión logística; deserción; ingresantes

CONSTRUCCIÓN DE UNA ESCALA PARA MEDIR LA PROPENSIÓN DE USO DE PLATAFORMAS DE FINANCIAMIENTO MASIVO EN ESTUDIANTES DE CIENCIAS ECONÓMICAS

MARIANA GUARDIOLA y NORMA PATRICIA CARO

FACULTAD DE CIENCIAS ECONÓMICAS, UNIVERSIDAD NACIONAL DE CÓRDOBA.

mariaganaguardiola@eco.unc.edu.ar

El crowdfunding o financiamiento masivo es una manera de conseguir recursos financieros a través de plataformas en Internet a fin de que un gran número de individuos realicen aportes a una causa, logrando en su conjunto una contribución relevante. Debido a que en la literatura no existe consenso sobre escalas que midan la propensión al uso de este tipo de financiamiento, se definió un constructo referido a la valoración de plataformas de crowdfunding para el financiamiento de proyectos de calidad. A través de un modelo de medida, en el marco de los modelos de ecuaciones estructurales, se estimaron las relaciones causales del comportamiento de los potenciales usuarios de crowdfunding y el efecto entre los indicadores propuestos y la variable latente. En el año 2016 se realizó una encuesta piloto a 377 estudiantes avanzados de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba (UNC), a fin de evaluar su propensión al uso de plataformas de financiamiento masivo para proyectos de calidad. En base a la literatura existente se definieron indicadores de satisfacción, reconocimiento, valoración, aporte y confianza en sí mismo y a este medio de financiamiento, obteniéndose relaciones significativas entre cada uno de los indicadores y la variable latente. El estadístico Chi cuadrado indicó que el modelo es adecuado ($p\text{-value}: 0,079$), al igual que las medidas de bondad de ajuste (RMSE del 0,05 y SRMR del 0,04). A partir de los resultados obtenidos se pretende extender el análisis a todos los estudiantes de grado de la UNC, considerando otras dimensiones de relevancia.

Palabras claves: ecuaciones estructurales; crowdfunding; variable latente

CUMPLIMIENTO DE LA LEY NACIONAL DE CONTROL DE TABACO EN PUNTOS DE VENTA CERCANOS A ESCUELAS SECUNDARIAS Y SU RELACIÓN CON EL CONSUMO DE TABACO EN ADOLESCENTES

ADRIANA PÉREZ¹, GERARDO CUETO², SANDRA BRAUN³ y RAÚL MEJÍA³

¹ FACULTAD DE CIENCIAS EXACTAS Y NATURALES (UBA) - CEDES. ² FACULTAD DE CIENCIAS EXACTAS Y NATURALES (UBA). ³ HOSPITAL DE CLÍNICAS JOSÉ DE SAN MARTÍN - CEDES.

adrianaperez000@gmail.com

Introducción: La promoción del tabaco aumenta el riesgo de que los jóvenes se inicien en el consumo de tabaco. Los puntos de venta de tabaco (PVT) representan un espacio de marketing de interés para la industria tabacalera. Objetivos: 1- Describir la cantidad y proximidad de PVT alrededor de escuelas secundarias. 2- Determinar el cumplimiento de la Ley Nacional de Control de Tabaco, que regula la promoción y publicidad de tabaco, en dichos PVT. 3- Determinar si el consumo de tabaco de los estudiantes se asocia a la cantidad, proximidad y grado de cumplimiento de la ley de los PVT cercanos a las escuelas a las que concurren. **Métodos:** Una encuesta autoadministrada fue aplicada a 3113 estudiantes de 33 escuelas secundarias públicas y privadas de Buenos Aires, Córdoba y Tucumán y se registró el consumo de tabaco. Se identificaron todos los PVT ubicados en un radio de 300 metros alrededor de las escuelas ($n=319$) y observadores entrenados determinaron el cumplimiento de siete artículos de la Ley Nacional de Control de Tabaco. Se aplicaron modelos de regresión multinivel para estudiar la relación entre ser fumador actual (últimos 30 días) o experimentador (fumó alguna vez pero no actualmente) en función de la cantidad y proximidad de los PVT y la cantidad de artículos de la ley incumplidos por los PVT. **Resultados:** Se encontraron entre 0 y 26 PVT alrededor de las escuelas (media=10,3, DE: 6,9). El 70% de las escuelas tuvo al menos un PVT a menos de 100 metros. El 99% de los PVT presentaron al menos una violación a alguno de los 7 artículos evaluados y el 45% incumplió al menos 4 artículos. Las violaciones más habituales fueron el exceso de tamaño y cantidad de letreros publicitarios de tabaco (63%) y la ausencia de advertencias sanitarias en los mismos (59%). El 17,7% de los estudiantes fue fumador actual y el 18,9% fue experimentador. No se encontró asociación significativa entre uso de tabaco en los estudiantes y las variables asociadas a PVT. **Conclusiones:** Si bien la Ley Nacional de Control de Tabaco fue promulgada en 2011 se observa un fuerte incumplimiento de la misma en áreas cercanas a escuelas secundarias, exponiendo a los jóvenes a mayor exposición de marketing de tabaco. Si bien nuestros resultados no muestran una mayor prevalencia de tabaquismo asociada a publicidad en PVT, no se puede descartar un efecto de la promoción de tabaco a mediano plazo.

Palabras claves: tabaquismo; adolescentes; modelos multinivel

DESARROLLO DE UN MODELO PREDICTIVO PARA EL ÍNDICE DE RETENCIÓN DE COMPUESTOS ORGÁNICOS VOLÁTILES

*RICARDO BAQUERO¹, ROMMEL LARGO¹, VIVIANA MARQUEZ¹, CRISTIAN
ROJAS², PIERCOSIMO TRIPALDI³ y MÓNICA BALZARINI⁴*

¹ MAESTRÍA EN ESTADÍSTICA APLICADA. FACULTAD DE CIENCIAS AGROPECUARIAS, FACULTAD DE CIENCIAS ECONÓMICAS Y FACULTAD DE MATEMÁTICA, ASTRONOMÍA, FÍSICA Y COMPUTACIÓN. UNIVERSIDAD NACIONAL DE CÓRDOBA, CÓRDOBA. ² INSTITUTO DE INVESTIGACIONES FISICOQUÍMICAS TEÓRICAS Y APLICADAS (INIFTA), CONICET, UNLP, DIAG. 113 Y 64, C.C. 16, SUCURSAL 4, 1900 LA PLATA, ARGENTINA. ³ VICERRECTORADO DE INVESTIGACIONES. UNIVERSIDAD DEL AZUAY, AV. 24 DE MAYO 7-77 Y HERNÁN MALO. CUENCA, ECUADOR. ⁴ FACULTAD DE CIENCIAS AGROPECUARIAS. UNIVERSIDAD NACIONAL DE CÓRDOBA. CONICET. AV. VALPARAISO SN CC 509. CIUDAD UNIVERSITARIA, CÓRDOBA, ARGENTINA.

crojasvilla@gmail.com

El objetivo de este trabajo es el desarrollo de una relación cuantitativa estructura-propiedad (QSPR) para predecir los índices de retención de Kovats de compuestos volátiles orgánicos. Se trabajó con 63 compuestos detectados mediante cromatografía de gases/espectrometría de masas en la pimienta negra (*Piper nigrum L.*). Para cada compuesto volátil se calcularon 1642 descriptores moleculares, independientes de la conformación, en el programa Dragon versión 7. La base de datos se dividió de forma aleatoria en grupos de calibración y validación con 25 moléculas cada uno y un grupo de predicción con 13 compuestos. Seguidamente, se usó un método secuencial de selección de variables sobre los 50 compuestos para la búsqueda de los mejores modelos de regresión de 1 a 7 descriptores. El criterio de selección fue la desviación estándar residual (S) durante validación. En cada regresión se han analizado el coeficiente R² y S como parámetros estadísticos de calidad. El mejor modelo QSPR quedó constituido por 3 descriptores moleculares (R² = 0.93 y S = 60), cuya calidad ha sido verificada mediante las técnicas de validación interna de dejar-uno-fuera (R² = 0.88 y S = 81.5) y dejar-varios-fuera de ventanas venecianas con 5 grupos de validación (R² = 0.88 y S = 76.6). Adicionalmente, se evaluó la capacidad predictiva del modelo QSPR mediante la predicción de los índices de retención de los 13 compuestos externos al ajuste del modelo. Finalmente, se ha definido el dominio de aplicabilidad mediante el cálculo del valor de influencia crítico ($h^* = 0.24$), con el cual se ha verificado que todos los índices de retención predichos pertenecen al dominio de ajuste del modelo. La relación cuantitativa estructura-propiedad estuvo constituida por los descriptores índice anillo distancia/detour de orden 4 (D/Dtr04), índice tipo Randi; promedio a partir de la matriz de distancias topológicas (ChiAD) y autocorrelación de Moran a desplazamiento 1 ponderado por el volumen de van der Waals (MATS1v).

Palabras claves: perfiles aromáticos; regresión; qspr

DIFERENCIACIÓN POR CALIDAD EN EL COMERCIO GLOBAL DE ALIMENTOS PESQUEROS: UN MODELO GRAVITACIONAL A NIVEL DE PRODUCTO

MARÍA VICTORIA LACAZE¹ y OSCAR MELO²

¹UNIVERSIDAD NACIONAL DE MAR DEL PLATA. ²PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE.

mvlacaze@mdp.edu.ar

Los esquemas de eco-etiquetado constituyen una importante herramienta de gestión de la calidad en el mercado global de alimentos pesqueros. Recientemente, se ha discutido si estos esquemas obstaculizarían el acceso al mercado por parte de productos procedentes de pesquerías (y países) del hemisferio sur. Sin embargo y hasta el presente, no hay análisis econométricos que brinden evidencia a favor del desempeño del eco-etiquetado como catalizador del o barrera al comercio de estos alimentos. Este trabajo contribuye a esa área de vacancia, mediante la estimación de un modelo gravitacional estructural, sobre un panel multidimensional de datos de comercio bilateral de alimentos pesqueros entre 38 países que, en términos de valor y para el período enero 2010–diciembre 2014, concentran el 80% de los flujos comerciales globales. La fuente principal de información es la base de exportaciones mensuales UN-Comtrade, a cuatro dígitos de desagregación del sistema armonizado (HS). Dado que el HS no efectúa desagregaciones por estándares de calidad, se realizó un matching entre dicha fuente y una base de datos auxiliar, para identificar los flujos comerciales que corresponden a productos eco-etiquetados. Esta base auxiliar fue elaborada a los fines del estudio y contiene información procedente de los sitios web de los principales esquemas de eco-etiquetado vigentes. Debido a la separabilidad del modelo gravitacional, la estimación se efectúa a nivel de producto. El tratamiento de los términos de resistencia multilateral se resuelve con la incorporación de efectos fijos direccionales, por país y período. La estimación en forma multiplicativa, mediante el estimador máximo verosímil Pseudo-Poisson, soluciona el sesgo por selección muestral que genera la alta prevalencia de ceros en la variable dependiente y aborda la heterosce-

dasticidad de los datos, circunstancias habituales en este tipo de estudios. La aproximación a los costos comerciales bilaterales incluye la distancia marítima entre los principales puertos pesqueros, la certificación mediante eco-etiqueta y otras covariables de uso tradicional, como la contigüidad, el idioma común, la relación colonial y la pertenencia a un mismo acuerdo comercial regional, además de los aranceles bilaterales. El modelo estocástico estimado queda expresado por: $X_{ij,t} = \exp[p_{ii,t} + n_j t + b_1 D_{ij,t} + b_2 C_{ij,t} + b_3 I_{ij,t} + b_4 C_{ij,t} + b_5 E_{ij,t} + b_6 A_{ij,t}] \times E_{ij,t}$. A partir de los coeficientes estimados, se obtuvieron los efectos de volumen o elasticidades parciales de las exportaciones, así como el efecto arancelario equivalente de la certificación mediante eco-etiqueta, es decir el arancel ad valorem cuya aplicación o remoción (según cada producto) equivale al efecto, estadísticamente significativo, que la eco-etiqueta produce en el comercio de estos alimentos.

Palabras claves: modelo gravitacional; comercio; alimentos pesqueros; diferenciación; eco-etiquetado; arancel equivalente

EL PROTOCOLO Y SU EFECTO SOBRE LA CONTINUIDAD DE LA EMPRESA FAMILIAR A TRAVÉS DE MODELOS DE ECUACIONES ESTRUCTURALES

MARÍA DE LOS ÁNGELES LUCERO BRINGAS¹ y NORMA PATRICIA CARO²

¹UNIVERSIDAD CATÓLICA CÓRDOBA. ²UNIVERSIDAD NACIONAL DE CÓRDOBA.

pacaro@eco.unc.edu.ar

Las empresas familiares son representativas de la economía en la provincia de Córdoba, ellas se caracterizan por una fuerte vocación de continuidad como parte del legado familiar. Sin embargo, su permanencia se haya obstaculizada en la mayoría de los casos por la falta de gestión de la familia empresaria. El alto índice de desaparición obedece en la mayoría de los casos a contrariedades de índole familiar. El protocolo familiar es una herramienta que aporta a la continuidad de las empresas familiares facilitando la gestión empresarial y de la familia empresaria. Se materializa a través de un acuerdo escrito donde los miembros de la familia se anticipan a los tópicos que pueden afectar la continuidad, previendo principios y normas cuyo cumplimiento estén orientados a una mayor unidad familiar y fortalecimiento de la empresa. El protocolo es un instrumento de gestión indispensable para la supervivencia de la empresa familiar y familia empresaria. Este trabajo tiene como objetivo, encontrar un análisis causal entre el comportamiento de las empresas familiares en relación al protocolo familiar como herramienta de gestión en pos a la continuidad del negocio familiar. La muestra está formada por 220 empresas familiares protocolizadas (16%) y no protocolizadas (84%), donde el 46,8% se encuentra en la segunda generación, un 32,7% aún está en primera y un 20,5% se encuentra en tercera o superior. Se aplicaron los modelos de ecuaciones estructurales (structural equations models SEM), donde por un lado, se define la variable latente "continuidad" que indica la percepción que tienen los miembros de las empresas familiares y por otro lado si poseer protocolo tiene implicancia en la continuidad. El Alpha de Cronbach indica que el instrumento es adecuado para lo que desea medir (0,92) y los p-value obtenidos por el SEM indican que las afirmaciones fueron significativas en la definición del constructo y confirmán la hipótesis de que invertir un protocolo familiar influye a la continuidad de la empresa familiar. Las medidas de bondad de ajuste: RMSEA (0,064) y SRMR (0,058) resultaron aceptables. Poseer un protocolo familiar que incluya los lineamientos para una gestión familiar y empresarial contribuye a la continuidad de la empresa familiar con todos los beneficios que ello implica tanto para la propia familia empresaria como para la sociedad y economía en la cual se encuentra inmersa.

Palabras claves: ecuaciones estructurales; alpha de cronbach; empresas familiares; protocolo

EPIDEMIOLOGÍA DE LA MALARIA EN PANAMÁ DESDE LA PERSPECTIVA DEL NIÑO OSCILACIÓN DEL SUR

LISBETH AMARILIS HURTADO A.¹, JOSÉ CALZADA¹, CHYSTRIE RIGG¹, MILIXA PEREA¹, SAHIR DUITARY¹, MILAGROS CASTILLO¹, ELSA BENAVIDES² y LUIS CHAVES³

¹INSTITUTO CONMEMORATIVO GORGAS DE ESTUDIOS DE LA SALUD. ²MINISTERIO DE SALUD DE PANAMÁ. ³PROGRAMA DE INVESTIGACIÓN EN ENFERMEDADES TROPICALES. ESCUELA DE MEDICINA VETERINARIA, UNIVERSIDAD NACIONAL, COSTA RICA..

lhurtado@gorgas.gob.pa

Introducción. En Panamá la malaria continúa siendo un reto para la salud pública; que con el tiempo ha ocupado una importante dimensión social de difícil intervención. En la epidemiología de la malaria se observa una tendencia descendente desde el 2005, cuando ocurrió la notificación de casos más alta (5095) y los últimos registros de mortalidad por esa enfermedad. En este lapso la transmisión de la malaria ha mostrado variaciones entre provincias y de un año a otro. Aún así, la morbilidad se mantiene focalizada y relativamente baja. Al igual que el resto de Centroamérica, la mayoría de las infecciones por malaria han sido por *Plasmodium vivax* (97 %). Geográficamente los casos se concentran en poblados indígenas. Varios estudios han sugerido que en años recientes el ambiente se ha comportado de manera anómala cuando se compara con las décadas pasadas. Esto indica que los programas de control de la malaria se enfrentarán ahora a una mayor incertidumbre en los patrones de la enfermedad. En consecuencia, saber hasta qué grado los factores ambientales posiblemente han cambiado patrones de transmisión e inhibido la probabilidad de éxito de Programas de Control es una incógnita importante. **Desarrollo.** Desde esta perspectiva, se realiza un estudio de tipo ecológico, en el cual se busca analizar la asociación entre los cambios ambientales y la transmisión espacio-temporal de la malaria en Panamá desde 1990 a 2015. Específicamente estudiamos los impactos de la variabilidad climática, medido por las diferentes fases del Niño Oscilación Sur (ENOS). Para ello, se aplicó un análisis de sincronía para estudiar los cambios en la incidencia de los casos de malaria entre las distintas provincias. **Resultados.** Los resultados sugieren que el Niño tiene un impacto sobre la epidemiología de la malaria en Panamá, sin embargo varía geográficamente. **Conclusión.** La última década ha sido testigo de los avances contra la eliminación de la malaria. La información epidemiológica muestra una reducción que vislumbra alcanzar esta meta. El éxito ha estado supeditado a lograr arraigar en los poblados indígenas el compromiso de eliminar la malaria. Hoy el escenario involucra, frenar los casos de *Plasmodium falciparum* que podrían deteriorar aún más la frágil condición de vida de los pueblos originarios de Panamá y establecer las medidas para reducir la vulnerabilidad de estas comunidades ante los cambios del ambiente.

Palabras claves: malaria; plasmodium vivax; enso; panamá

ESTIMACION DEL COMPORTAMIENTO ASIMETRICO DE LOS COSTOS CON UN MODELO LINEAL MIXTO COMPARACION DE EMPRESAS ARGENTINAS, CHILENAS, PERUANAS Y COLOMBIANAS

MARIA INES STIMOLO y MAXIMILIANO IGLESIAS

FACULTAD DE CIENCIAS ECONOMICAS.

mstimolo@eco.unc.edu.ar

En este trabajo se propone un modelo mixto para cuantificar el incremento (o disminución) de los gastos por cada incremento (o disminución) porcentual de los ingresos por venta de las empresas

extendiendo el modelo de regresión log lineal por partes propuesto por Anderson et.al. (2003). Con este modelo se verifica empíricamente la falta del supuesto de proporcionalidad simétrica entre los costos y el nivel de ventas, sobre el cual se basan los modelos de costos a corto plazo y los métodos de costeo. El modelo lineal mixto permite predecir el comportamiento para cada empresa cuando los coeficientes aleatorios son significativos. El análisis se realiza sobre una muestra de empresas argentinas, chilenas, peruanas y colombianas que hacen oferta pública de sus acciones durante el período 2004-2012. El modelo de comportamiento asimétrico de los costos está basado en la teoría de las decisiones gerenciales en relación al ajuste de los costos de las empresas, siendo estas decisiones afectadas por, las expectativas económicas, entre otros. La empresas analizadas corresponden a países con ambientes y expectativas macroeconómicos diferentes. Además, todas estas economías fueron de una manera u otra afectadas por la crisis del año 2008 por lo que se incorporan en el análisis los períodos 2004-2007 y 2008-2012, encontrando diferencias en el comportamiento de los costos de las empresas relacionados con el contexto económico de cada uno. Las empresas se clasificaron en sectores, los que fueron incorporados al modelo como una manera de incorporar estructura de costos diferentes. Los sectores tienen características diferentes en los países analizados por lo que el análisis comparativo resulta más complejo.

Palabras claves: datos longitudinales; modelos mixtos; costos asimétricos; economías diferentes

ESTRATEGIAS DE ANÁLISIS CON MODELOS LINEALES GENERALIZADOS MIXTOS PARA PREDICCIÓN DE NITRÓGENO MINERALIZADO APARENTE EN CULTIVO DE MAÍZ EN LA PROVINCIA DE BUENOS AIRES

*OLGA SUSANA FILIPPINI¹, VANINA MARIANA DELFINO² y MERCEDES
ZUBILLAGA³*

¹UNLU CS.BÁSICAS - FAUBA. ²UNTREF. ³FAUBA - FERTILIDAD.

sfilippini@unlu.edu.ar

Con el objeto de ajustar modelos predictivos del Nitrógeno mineralizado aparente (Nap), durante el ciclo del cultivo, a partir de variables climáticas, edáficas y espetrales a escala de parcela, se llevaron a cabo diversas estrategias de análisis Estadístico. La información de las variables edáficas se obtuvo a partir de un muestreo aleatorio de las parcelas cultivadas con maíz. Se consolida la base de datos, aplicando Modelos que puedan predecir la Mineralización del Nitrógeno evaluado en suelos argentinos de la planicie de la provincia de Buenos Aires, durante el ciclo del crecimiento del maíz. El tamaño de la muestra es de 139 datos, extraídos durante 6 campañas, considerando en este caso los años de cada campaña como el factor aleatorio, en 6 parcelas de producción de donde se escogieron, microparcelas para evaluar la mineralización del Nitrógeno. Como co-variables, se consideraron variables climáticas como las precipitaciones en distintos períodos y las temperaturas. Las variables edáficas incluidas en el modelo fueron, arena, arcilla y limo y la variable topográfica (loma, media loma o bajo) fue considerada factor fijo. Algunas variables edáficas y la topográfica resultaron estadísticamente significativas. Se comparan las posiciones topográficas, que representan el tipo de ambiente, al momento de medir la cantidad de nitrógeno mineralizado en la parcela, detectando diferencias significativas. Las predicciones del Nitrógeno Mineralizado aparente a partir de la selección de alguna de las variables relevadas y de Modelos Lineales Generalizados con diferentes funciones de enlace permite estudiar la variabilidad espacial y utilizar las mismas en las recomendaciones de fertilización con Nitrógeno en la agricultura de precisión.

Palabras claves: modelo lineal genera; nitrógeno aparente; funciones de enlace

ESTUDIO DE LOS “NI-NI” DEL GRAN CÓRDOBA A TRAVÉS DE UN MODELO DE REGRESIÓN LOGÍSTICA

NANCY STANECKA, MACARENA CANCINOS y PATRICIO CANALIS

FAC. DE CIENCIAS ECONÓMICAS. UNC.

nstanec@gmail.com

Si bien la trayectoria juvenil hacia la adultez se observa cada vez más heterogénea y compleja, es esperable que en esta etapa de la vida los jóvenes transitén por ciertos eventos claves que incluyen, entre otros, la finalización de los estudios y/o el ingreso al ámbito laboral. No obstante, los registros censales permiten afirmar que en el año 2010 el 16,7% de los habitantes de la provincia de Córdoba, que tenían entre 15 y 29 años, no trabajaba ni estudiaba. Este fenómeno, de jóvenes que ni trabajan ni estudian, es conocido con el nombre de la generación de los “ni-ni”, pudiendo pensarse la situación como de alto riesgo social. En este marco, el siguiente estudio presenta una caracterización general de los jóvenes e indaga sobre factores socio-demográficos que inciden en la condición de “ni-ni” de aquellos que habitan el aglomerado urbano Gran Córdoba. Para ello se recurre a la información que brinda el Censo Nacional de Población, Hogares y Vivienda 2010 y la Encuesta Permanente de Hogares (EPH) del último semestre del mismo año. En una primera etapa se realiza un análisis descriptivo de la población en estudio y posteriormente se utiliza un modelo de regresión logística binaria donde la variable respuesta indica si el individuo presenta el status “ni-ni”, considerando como variables predictoras a factores socio-demográficos correspondientes al joven y socio-económicos del hogar en el que el mismo habita. Este último análisis nos permite identificar como variables estadísticamente significativas en la condición de “ni-ni”, al sexo del joven, su nivel educativo y ciertas necesidades básicas insatisfechas del hogar donde él vive, entre otras.

Palabras claves: regresión logística; joven urbano; jóvenes “ni-ni”

ESTUDIO DEL COMPORTAMIENTO DE LOS ESTIMADORES DE REGRESIÓN EN GRANDES DIMENSIONES DE DATOS

IVÁN MILLANES, MARÍA BELÉN ALLASIA y MARTA BEATRIZ QUAGLINO

INSTITUTO DE INVESTIGACIONES TEÓRICAS Y APLICADAS DE LA ESCUELA DE ESTADÍSTICA (IITAE). FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA, UNR..

ivanmillanes26@gmail.com

Para estimar los parámetros de un modelo de regresión lineal, es común utilizar el método de mínimos cuadrados ordinarios (MCO), método popular por su fácil aplicación y buenas propiedades. Sin embargo, cuando el número de variables explicativas (p) es mayor que el número de observaciones (n) el estimador MCO no es único y pierde sentido la interpretación de los resultados. Incluso cuando p es menor pero cercano a n -pese a la existencia de una estimación mínimo-cuadrática única- puede no ser conveniente la utilización de este método debido a la gran variabilidad de las estimaciones. En este contexto es común utilizar métodos de regularización, que minimizan una Suma de Cuadrados del Error Penalizada: regresiones Ridge y LASSO. En el presente trabajo se efectuó un estudio por simulación para comparar el desempeño de los métodos de regresión en escenarios donde p es igual a n modificando el número de parámetros significativos en el modelo (k), esto es la cantidad de parámetros distintos de cero. A partir de la generación de conjuntos de observaciones de la variable respuesta (Y), se calculó el estimador MCO y se ajustaron los caminos de solución Ridge y LASSO para 50 valores del parámetro de penalización, reteniendo en cada caso

sólo el estimador del vector de parámetros que produce el menor Error Cuadrático Medio (ECM). Se construyeron las distribuciones empíricas de los estimadores obtenidos con cada método considerando globalmente aquellos que son significativos y los que no lo son. Se observó que para estimar parámetros no significativos tanto los estimadores Ridge como LASSO resultaron sesgados, siendo el método LASSO más preciso. Al estimar parámetros significativos y para k pequeño, ambas técnicas proveen resultados sesgados, siendo el desempeño de Ridge peor que el de LASSO en este sentido. Al aumentar k , el método LASSO fuerza a muchos estimadores a ser cero a pesar de que sean significativos, mientras que Ridge mantiene prácticamente el mismo comportamiento que para k pequeño. Por otra parte, MCO provee estimadores insesgados para todos los parámetros pero con variabilidad mucho mayor que las regresiones penalizadas. Al comparar los ECM promedios, se observó que la regresión LASSO tiene mejor desempeño que Ridge cuando k es pequeño. A medida que aumenta k , las regresiones penalizadas empeoran su performance, funcionando mejor Ridge que LASSO. En general el ECM promedio del método MCO fue significativamente mayor, por lo cual MCO no es una alternativa admisible en estos escenarios.

Palabras claves: mco; ridge; lasso; big data; ecm

EVALUACIÓN DE FACTORES INFLUYENTES EN LA PRESIÓN ARTERIAL EN PACIENTES HIPERTENSOS UTILIZANDO UN MODELO LINEAL MIXTO

*M. DEL CARMEN GARCÍA¹, NOELIA CASTELLANA¹, MARA CATALANO¹,
CECILIA RAPELLI¹ y CAROLINA CHACÓN²*

¹INSTITUTO DE INVESTIGACIONES TEÓRICAS Y APLICADAS. ESCUELA DE ESTADÍSTICA / FACULTAD DE CS ECONÓMICAS Y ESTADÍSTICAS (UNR), ARGENTINA. ²FUNDACIÓN ECLA, ARGENTINA.

mgarcia@fcecon.unr.edu.ar

La hipertensión arterial, que consiste en el aumento de la presión arterial, es una de las principales causas de las enfermedades cardiovasculares en el mundo y sus complicaciones son responsables de millones de muertes anuales. A pesar de que la hipertensión arterial es un factor de riesgo modificable y que existe una gran cantidad de medicamentos efectivos para su tratamiento, esta enfermedad crónica sigue teniendo una alta prevalencia asociada a altas tasas de desconocimiento y a un mal control de los valores de presión arterial en el tiempo (valores por encima de los óptimos). Esta falta de control adecuado obedece a diversas razones, entre las que se destacan los hábitos poco saludables y la mala adherencia terapéutica, es decir, el paciente no cumple con el tratamiento indicado. El objetivo de este trabajo es evaluar el efecto del grado de adherencia al tratamiento farmacológico y de otras variables sobre la presión arterial sistólica de un grupo de pacientes hipertensos que forman parte de un programa de atención y control de la hipertensión iniciado en el año 2014 en Rosario. Este programa contempló un seguimiento exhaustivo de los mismos en el tiempo, registrándose tanto características basales como los valores de presión arterial en cada una de las visitas realizadas. A partir del año 2016, con la incorporación del cuestionario validado de Morisky-Green, se comenzó a medir la adherencia al tratamiento farmacológico. Como consecuencia de esto, los pacientes que ya participaban del programa no cuentan con esta información previo a esta fecha. Una herramienta valiosa para el análisis de este tipo de datos son los modelos lineales mixtos que permiten modelar las múltiples mediciones de la presión arterial sistólica en función de las covariables y la correlación entre las mismas. El análisis de los resultados obtenidos permite evidenciar que el valor de la presión arterial sistólica basal, el tipo de tratamiento, el sexo del paciente y el grado de adherencia al tratamiento farmacológico influyen en los valores medios de la presión arterial.

Palabras claves: hipertensión arteria; adherencia al tratam; modelo lineal mixto; elección de co-variaab

EVALUACIÓN DE LA EFICIENCIA REPRODUCTIVA EN VACAS LECHERAS CON MODELOS DE COX MIXTOS

MONICA PICCARDI¹, GABRIEL BÓ² y MONICA BALZARINI¹

¹FACULTAD DE CIENCIAS AGROPECUARIAS – UNC; CONICET. ²INSTITUTO DE REPRODUCCIÓN ANIMAL (IRAC); INSTITUTO DE CIENCIAS BÁSICAS Y APLICADAS – UNVM.

monicapiccardi@gmail.com

Los establecimientos lecheros utilizan distintos indicadores para monitorear el desempeño reproductivo, como son los días vacíos (DV). Los DV son los días que transcurren desde que la vaca tiene un parto hasta que se preña nuevamente de manera efectiva. Existen diferentes herramientas metodológicas que pueden ser usadas para construir este indicador. El uso de medidas resúmenes como estadísticas descriptivas simples para estimar los DV a partir de los datos individuales de un rodeo puede introducir sesgos en la evaluación del desempeño reproductivo ya que éstas estadísticas sólo incluyen la información del tiempo en días desde parto a la concepción de vacas que se han preñado. Los animales que son descartados por no preñarse o por haber sufrido un aborto, o bien por haber sido descartados antes de tener la oportunidad de preñarse no aportan información para el cálculo descripto de los DV. Una mejor alternativa para la obtención de este indicador es su estimación a partir del análisis de sobrevida para el tiempo que transcurre hasta la preñez considerando todas las vacas. Esta técnica estadística no sólo incluye a los animales que experimentaron el evento, sino que también incluye a aquellos que no experimentan el evento o se pierden durante el periodo de observación. Los DV dependen del efecto del animal, pero también de la época del parto (época calurosa versus fresca) y del número de lactancia y en general del manejo reproductivo del tambo. El objetivo de este trabajo fue ajustar y comparar 6 modelos candidatos de regresiones de cox de efectos mixtos para modelar los DV. Para esto se utilizó una base de datos compuesta por 22.456 vacas de 35 tambos de la zona centro y sur de Santa Fe y provincia de Córdoba. Los efectos fijos del modelo con más parámetros fueron estación de parto, número de lactancia y su interacción; y los efectos aleatorios fueron el tambo y las vacas dentro de tambo. Usando la prueba de cociente de verosimilitud ajustada (LRT) se seleccionó el modelo donde el efecto aleatorio era las vacas independientemente de los tambos y ambos efectos fijos sin interacción. Los blup del efecto vaca fueron extraídos y usados para seleccionar a las mismas por fertilidad.

Palabras claves: días vacíos; efectos aleatorios; análisis de sobrevida

EVALUACIÓN DE LA EYACULACIÓN EN LESIONADOS MEDULARES TRAUMÁTICOS MEDIANTE ESTADÍSTICA MULTIVARIADA Y REGRESIÓN

LAURA CACHEIRO y ALBERTO RODRIGUEZ VELEZ

IREP/GCBA.

laura.cacheiro@gmail.com

RESUMEN La eyaculación está definida fisiológicamente como la expulsión enérgica y rítmica del semen hacia el meato uretral, ocurre como resultado de un reflejo medular que involucra la indemnidad de los segmentos medulares de Torácico 10 a Sacro 4, controlado por el Sistema Nervioso Central supramedular. Sin embargo la eyaculación es posible aun en ausencia de estas conexiones superiores por la existencia de un centro espinal generador reflejo de la eyaculación localizado en la región toracolumbar entre Toracico 10 – Lumbar 3. La lesión medular implica la pérdida total o parcial de las funciones motoras, sensitivas y autonómicas, por debajo del nivel de lesión. Luego de la lesión medular la mayoría de los hombres no pueden alcanzar una eyaculación durante

la masturbación o el coito. Existe amplia controversia respecto a las posibilidades eyaculatorias en pacientes con lesión medular, y muy poca investigación al respecto, dada la importancia que tiene, es necesario conocer cuales variables están implicadas en la presencia de eyaculación. Se desea conocer cuáles son las características de la lesión medular más fuertemente vinculadas con una eyaculación exitosa con o sin asistencia médica. Además de obtener información específica sobre la función sexual de los hombres con lesiones de la médula espinal. Para ello se registraron los siguientes datos de 185 pacientes: Nivel de lesión motora, Nivel de lesión sensitiva, erección psicogénica, erección refleja, contracción de esfínter anal y Escala ASIA. La aplicación de metodología estadística multivariada como, Análisis de Correspondencias múltiple, Regresión Logística y el manejo de datos faltantes permitió en primera instancia tener una idea del comportamiento de todas las variables en conjunto y luego ayudo a entender variables mas asociadas con la Eyaculacion. Este estudio confirma los hallazgos anteriores que la capacidad de alcanzar la eyaculación es deficiente en los hombres con lesión de la médula espinal. La discapacidad locomotora más severa podría afectar negativamente la vida sexual de las personas con lesión medular. Estas conclusiones se traducen en poder predecir la capacidad eyaculatoria y en consecuencia poder emprender acciones que permitan a los pacientes acceder a la paternidad y una mejor en su calidad de vida sexual.

Palabras claves: lesión medular; eyaculacion; descriptiva multivaria; regresion logistica

EVALUACIÓN DEL CRECIMIENTO DE POLLOS PARA CARNE A PARTIR DEL ANÁLISIS CONJUNTO DE DOS VARIABLES MEDIANTE UN MODELO NO LINEAL MULTIVARIADO

CECILIA RAPELLI¹, RICARDO DI MASSO² y MARÍA DEL CARMEN GARCIA³

¹ INSTITUTO DE INVESTIGACIONES TEÓRICAS Y APLICADAS DE LA ESCUELA DE ESTADÍSTICA. FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA. UNIVERSIDAD NACIONAL DE ROSARIO. ²INSTITUTO DE GENÉTICA EXPERIMENTAL. FACULTAD DE CIENCIAS MÉDICAS. UNIVERSIDAD NACIONAL DE ROSARIO. CIC-UNR. ³INSTITUTO DE INVESTIGACIONES TEÓRICAS Y APLICADAS DE LA ESCUELA DE ESTADÍSTICA. FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA. UNIVERSIDAD NACIONAL DE ROSARIO. CIC-UNR.

cecirapelli@hotmail.com

En las investigaciones biológicas es de frecuente interés describir el proceso de crecimiento de los integrantes de una población mediante un modelo estadístico que explique la modificación de un indicador dado en función de la edad cronológica. Para caracterizar ese proceso se efectúan mediciones en forma repetida de dicho indicador. Los datos longitudinales multivariados surgen cuando un conjunto de diferentes respuestas se mide repetidamente en el tiempo sobre una misma unidad. Resulta de interés para tales datos conocer cómo la evolución de una respuesta está relacionada con la evolución de otra respuesta y/o como la asociación entre las distintas respuestas evoluciona con el tiempo. Por otro lado, la dinámica del crecimiento puede ser descripta mediante curvas de crecimiento sigmoideas. Los modelos no lineales mixtos multivariados proveen una herramienta útil para analizar este tipo de datos en los cuales la relación entre las variables explicativas y las variables respuesta puede modelarse con una función no lineal, permitiendo a los parámetros diferir entre las unidades. En la Cátedra de Genética de la Facultad de Ciencias Veterinaria se llevó a cabo un estudio con pollos destinados a la producción de carne con el fin de estudiar el comportamiento del peso corporal –estimador de la biomasa sustentada- y la longitud de la caña –estimador de la base ósea de sustentación- en función de la edad cronológica. Para esto se registraron ambas variables a intervalos semanales entre el nacimiento y la faena a las 12 semanas de edad. El estudio de la modificación del peso en función de la modificación de la longitud de la caña tiene como particularidad que, dado que la caña tiene un crecimiento finito, la máxima longitud que puede alcanzar limita el rango de variación del peso corporal. Se ajustó un modelo no lineal mixto multivariado utilizando la curva de Gompertz asumiendo que sus tres parámetros varían estocásticamente entre las unidades.

El análisis de los resultados obtenidos permite concluir que la tasa de crecimiento de la longitud de la caña y el peso máximo alcanzado por los pollos están negativamente correlacionados (-0,45) indicando que cuanto más rápido crezca la caña menor será el peso máximo. Se observa, además, una fuerte correlación (0,75) entre el comportamiento de la longitud de la caña y el correspondiente al peso corporal de los pollos en función del tiempo, consistente con la condición de indicadores de crecimiento dimensional de ambas variables.

Palabras claves: datos longitudinales; curva crecimiento; modelos no lineales

EVALUACIÓN DEL EFECTO DE LA CIRUGÍA DE PTERIGIÓN EN LA EVOLUCIÓN DEL ASTIGMATISMO CORNEAL A TRAVÉS DEL USO DE MODELOS LINEALES GENERALES MIXTOS

MARCOS PRUNELLO¹, EZEQUIEL ARANDA², JÉSICA DIMATTIA², RODRIGO MANTARAS², ELINA CIARLO², BRUNO ZEGA², LUCIANA MICHELETTI² y NADIA SICA²

¹ÁREA ESTADÍSTICA Y PROCESAMIENTO DE DATOS. FACULTAD DE CIENCIAS BIOQUÍMICAS Y FARMACÉUTICAS, UNIVERSIDAD NACIONAL DE ROSARIO. ²CÁTEDRA DE OFTALMOLOGÍA DE LA UNIVERSIDAD NACIONAL DE ROSARIO, HOSPITAL PROVINCIAL DEL CENTENARIO.

mprunello@fbioyf.unr.edu.ar

El pterigión es un crecimiento fibrovascular del tejido conectivo conjuntival, localizado en forma de ala sobre la córnea. Puede causar aplanamiento de la córnea e inducir astigmatismo. El objetivo de este trabajo es evaluar el efecto de la excéresis del pterigión en el astigmatismo corneal. Se evaluaron de manera prospectiva 102 pacientes que asistieron a la consulta al Servicio de Oftalmología del Hospital provincial del Centenario de Rosario, y que fueron sometidos a cirugía de pterigión según la técnica de injerto conjuntival libre entre agosto de 2014 y agosto de 2016. Se registró la edad, sexo, ubicación de la degeneración (nasal o temporal), aspecto biomicroscópico (rosado, hiperémico o blanco-atrófico) y clasificación en grados de acuerdo a su extensión sobre la córnea. Todos los pacientes incluidos en el estudio fueron sometidos a un examen oftalmológico completo incluyendo agudeza visual y topografía corneal para cuantificar el nivel de astigmatismo en la primera consulta, y luego de uno y tres meses de la cirugía. La evolución en el tiempo del astigmatismo fue estudiada con un modelo lineal general mixto para datos longitudinales. El modelo incluyó efectos fijos para el grado (1, 2, 3 o 4), el aspecto (blanco, rosado o hiperémico), la lateralidad (ojo izquierdo o derecho) y el tipo (sólo nasal/temporal o ambos), y efectos aleatorios para la identificación del paciente. Se encontró que las variables aspecto, lateralidad y tipo del pterigión no influyen significativamente en los valores medios de astigmatismo ($p = 0.1966$, $p = 0.6306$ y $p = 0.5975$, respectivamente). Por el contrario, los efectos del tiempo, del grado y de la interacción entre ambos resultaron estadísticamente significativos ($p = <0.0001$ en los tres casos). Para los ojos con el grado más leve, no se observan diferencias significativas en el promedio de astigmatismo a través de los tres controles. En cada uno de los restantes grados, los valores promedios en las dos evaluaciones post-operatorias no difieren entre sí, pero son significativamente menores que al promedio del pre-operatorio. Finalmente, 49 casos (48.0%) evidenciaron una mejora en la agudeza visual, 47 (46.1%) permanecieron igual y 6 (5.9%) empeoraron, siendo estos casos con recidivas o con grado inicial I o 2. Este estudio corrobora que la cirugía exitosa de excéresis de pterigión reduce el astigmatismo inducido por el pterigión y mejora la agudeza visual, siendo el efecto beneficioso más notable en los pacientes con pterigión de grados más avanzados.

Palabras claves: pterigión; astigmatismo; modelo lineal genera; oftalmología; cirugía

METODOLOGÍA PARA LA PREDICCIÓN DEL COEFICIENTE DE RETENCIÓN DE UN HERBICIDA EN SUELO

FRANCA GIANNINI KURINA¹, SUSANA HANG², ARIEL RAMPOLDI² y MÓNICA BALZARINI¹

¹FACULTAD DE CIENCIAS AGROPECUARIAS, UNIVERSIDAD NACIONAL DE CÓRDOBA. CONICET. ²FACULTAD DE CIENCIAS AGROPECUARIAS, UNIVERSIDAD NACIONAL DE CÓRDOBA.

francagianninikurina@gmail.com

La dinámica de un fitosanitario en suelo es un proceso complejo que depende de propiedades edafoclimáticas del sitio donde se aplica. Uno de los principales mecanismos que caracterizan esta dinámica es la retención parametrizada por el coeficiente de adsorción (Kd), relación entre lo que queda retenido en el suelo y lo que se deriva. El objetivo de este trabajo fue generar un modelo predictivo del Kd de glifosato para suelos de la provincia de Córdoba. Se calculó el Kd en 90 muestras de suelo distribuidas en el territorio y seleccionadas según un HLC basado en variables edafoclimáticas en búsqueda de variabilidad de las co-variables a usar en la modelación. El Kd de glifosato se distribuyó como una gama (0,035; 1,129). Mediante árboles de regresión mejorados por re-muestreo (Boosting RT) aplicados sobre 20 co-variables edafoclimáticas y Kd como variables dependiente, se seleccionaron las variables de mayor capacidad contribución en la explicación de la variabilidad del Kd. Se ajustó un modelo lineal generalizado de regresión múltiple, con distribución gama y enlace log y un modelo lineal al Logaritmo del Kd. El Error Cuadrático Medio de Predicción relativo a la media, del modelo en escala logarítmica fue del 7%. Se priorizó este modelo sobre el modelo lineal generalizado por su simplicidad. Los resultados indicaron que los efectos del pH del suelo y la textura fueron las propiedades edáficas de mayor relevancia para predecir la retención de glifosato en los suelos muestreados. El impacto del Al fue modelado con un polinomio de segundo orden, mientras que para pH, Arena y Limo se usaron términos de primero orden. La ecuación resultante fue usada como valor esperado del Kd con la finalidad de mapear e identificar áreas de mayor vulnerabilidad a la contaminación por este herbicida.

Palabras claves: árboles de regresión; glifosato; suelos; predicción

MODELOS ADITIVOS GENERALIZADOS UNA APLICACIÓN EN EL CAMPO DEL MARKETING DIRECTO

RODRIGO LÓPEZ y FERNANDA MENDEZ

ITAE. FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA. UNIVERSIDAD NACIONAL DE ROSARIO.

fmendez@fcecon.unr.edu.ar

Cada vez más empresas utilizan modelos predictivos para optimizar sus campañas de marketing directo; estos modelos, usualmente conocidos como modelos de propensión de compra, permiten estimar la probabilidad que tiene una persona de aceptar una determinada oferta, en base a un conjunto de características particulares. A través de los modelos de propensión de compra las empresas pueden orientar sus ofertas hacia aquellos individuos con mayor probabilidad de aceptación, incrementando las tasas de respuesta de las campañas y disminuyendo los costos asociados a las mismas. Tradicionalmente, los modelos de regresión logística han sido los más utilizados para modelar resultados binarios como es el caso de la respuesta a una campaña de marketing. Estos modelos poseen importantes virtudes: son capaces de combinar simplicidad con buena performance, son

fáciles de interpretar y relativamente robustos. Los modelos de regresión logística imponen una forma paramétrica (muchas veces lineal) para el efecto de cada variable predictora, por lo que han sido criticados por fallar en la captación de relaciones no lineales más complejas que a menudo se presentan en la práctica, conduciendo en algunos casos a resultados no satisfactorios. Como alternativa, los modelos aditivos generalizados (GAM), introducidos por Hastie y Tibshirani (1986, 1990), omiten supuestos sobre la forma en que los predictores se relacionan con la respuesta y de esta manera proveen la capacidad de detectar relaciones más o menos complejas que de otro modo se perderían. Estos modelos utilizan funciones no paramétricas para describir el efecto de cada predictor, permitiendo un ajuste más flexible que el que se puede obtener con los modelos de regresión logística. En este trabajo se presentan los modelos aditivos generalizados y su aplicación en el campo del marketing directo. Se utilizan datos reales de una campaña de marketing de un reconocido banco nacional, con el objetivo de mostrar la utilidad de los GAM como alternativa y/o complemento de otros métodos tradicionales como la regresión logística. La comparación realizada entre los GAM y los modelos de regresión logística no pretende ser una comparación exhaustiva entre las dos metodologías. Han quedado sin considerar características importantes tales como robustez, estabilidad, tiempos de procesamiento, etc., que podrán incluirse en posteriores estudios. Del mismo modo sucede con ciertos aspectos propios de los GAM, han quedado pendientes análisis más profundos sobre la selección de los parámetros de suavizado del modelo así como la sensibilidad de los GAM ante la elección de distintos métodos de suavizado.

Palabras claves: modelo aditivo gener; suavizado; regresión logística; marketing directo

MODELOS DE CRECIMIENTO PARA GOMPHRENA PULCHELLA Y GOMPHRENA PERENNIS

ERICA RAÑA¹, MARÍA DEL C OCHOA², SALVERA PRIETO ANGUEIRA³ y
SALVADOR CHAILA²

¹CÁTEDRA DE BIOESTADÍSTICA Y DISEÑO-FACULTAD DE AGRONOMÍA Y AGROINDUSTRIAS. UNIVERSIDAD NACIONAL DE SANTIAGO DEL ESTERO. ²CÁTEDRA DE MATOLOGÍA - FACULTAD DE AGRONOMÍA Y AGROINDUSTRIAS. UNIVERSIDAD NACIONAL DE SANTIAGO DEL ESTERO. ³CÁTEDRA DE AGROCLIMATOLOGÍA - FACULTAD DE AGRONOMÍA Y AGROINDUSTRIAS. UNIVERSIDAD NACIONAL DE SANTIAGO DEL ESTERO.

erica.ra511@gmail.com

Modelos de crecimiento para Gomphrena pulchella y Gomphrena perennis. Raña,1 E. Ochoa,M. del C.2 Prieto Angueira, S.3 y Chaila, S.2 erica.ra511@gmail.com El propósito de éste trabajo fue encontrar una función que permita estimar la biomasa (expresada como peso seco en g.pl-1) en función de unidades térmicas (UT) de dos malezas: Gomphrena pulchella y Gomphrena perennis. El peso seco fue determinado en distintos momentos del crecimiento de plantas de G. pulchella emergidas el 19/01/14 y cosechada el 20/05/14 y de plantas de G. perennis emergidas el 02/12/15 y cosechadas el 03/03/16. El conocimiento de dicha función es importante porque permite explicar biológicamente la asignación de la biomasa por planta en función de la acumulación de temperatura y la modificación de las etapas vegetativas y reproductivas entre especies y años. Se ajustaron modelos no lineales para el Peso Seco de G. pulchella y G. perennis usando el módulo de Modelos no lineales Mixtos de Infostat v. 2017. Para corregir la falta de homogeneidad de varianzas, se probaron modelos con distintas funciones de varianza: VarPower, que la relaciona con la media a través de una función de potencia, Var ConstPower, similar a la anterior, pero corrida por una constante y Var Exp, una función exponencial. El modelo adecuado se seleccionó en base al criterio de información de Akaike (AIC), y teniendo en cuenta también el error estándar relativo de los estimadores (error estándar dividido por la estimación del parámetro). En el caso de G. pulchella se seleccionó el modelo de Gompertz para expresar acumulación de materia seca (PS) en función de UT, con una función de varianza Var Power, que resultó la mas adecuada para corregir la heterocedasticidad. $PS = \alpha * exp(-\beta * exp(-\gamma * UT)) = 21.46 * exp(-28.09 * exp(-0.00028 * UT))$ Siendo PS: peso seco de la biomasa; α : el valor máximo que alcanza la biomasa; β : número positivo que

desplaza el modelo de izquierda a derecha, gamma: tasa de aumento de la biomasa y UT unidades térmicas. En el caso de *G.perennis*, el modelo seleccionado es el logístico con función de varianza exponencial. $PS=\alpha/(1+\exp(-(UT-\beta)/\gamma))$ $PS=39.52/(1+\exp(-(UT-1325.95)/198.04))$ Siendo PS: peso seco de la biomasa; α : el valor máximo que alcanza la biomasa; β : número positivo, γ : tasa de aumento de la biomasa y UT unidades térmicas.

Palabras claves: modelos; crecimiento; gomphrena

MODELOS ESTADÍSTICOS INFLADOS EN CERO PARA ANALIZAR DATOS BIOLÓGICOS: MICROCAVIA AUSTRALIS COMO CASO DE ESTUDIO

CLAUDIA DE LOS RÍOS y NATALIA ANDINO

DEPARTAMENTO DE BIOLOGÍA. FACULTAD DE CIENCIAS EXACTAS FÍSICAS Y NATURALES. UNIVERSIDAD NACIONAL DE SAN JUAN.

cfdelorios@unsj-cuim.edu.ar

Tradicionalmente, en la ciencias biológicas, los datos comportamentales se analizaban utilizando ANOVA. Sin embargo, las variables respuesta que resulta de un proceso de conteo, tales como la frecuencia de un comportamiento particular, son mejor modeladas usando distribuciones de probabilidad discretas tales como la distribución de Poisson siempre que sea posible. Algunas veces al usar esta distribución Poisson se presenta sobredispersión. En este caso, el término de error debe incluirse en el modelo dando lugar a un Modelo Binomial Negativo. Sin embargo, en ecología comportamental una característica distintiva de estos sets de datos comportamentales es su tendencia a contener una gran cantidad de ceros, por lo que se ha sugerido que el uso de Modelos Inflados en cero es el mejor enfoque para este tipo de datos ya que permite analizar datos con exceso de ceros. El objetivo de este estudio es realizar un trabajo comparativo desarrollando modelos tradicionales y modelos inflados en cero que permitan describir mejor datos biológicos en ecología comportamental, y que puedan utilizarse para realizar inferencias y predicciones de eventos futuros. Se trabajó con datos de un proyecto de ecología comportamental de un roedor social (*Microcavia australis*) que habita la provincia de San Juan. El proyecto busca conocer si las variables sociales (número de machos o número de hembras) o la época de reproducción influyen en el comportamiento agonístico de los machos de esta especie. Para esto se contó con datos de frecuencia de dicho comportamiento. Los datos presentaban un exceso de ceros (78%) y se analizaron considerando distribución Gaussiana, Poisson, Binomial Negativa, Inflado Cero Poisson e Inflado Cero Binomial Negativa. Los resultados mostraron que para este conjunto de datos los Modelos Inflados en Cero obtuvieron mejores resultados que los modelos tradicionales. Esto fue evidente a partir de los valores de AIC y de gráficos de ajuste de los residuos. El Modelo Inflado en Cero muestra que el número de hembras y la época reproductiva son las variables que mejor explican los comportamientos agonísticos en este roedor. Dado que los resultados de los Modelos Inflados en Cero proporcionaron una resolución estadística que correspondía con las respuestas biológicas esperadas, se sugiere que este análisis es una opción óptima para este tipo de datos.

Palabras claves: ecología comportamen; modelos inflados en; microcavia australis

MODELOS PARA EL TRATAMIENTO DE SESGOS EN ESTUDIOS DE DATOS AGREGADOS EN ESPACIO Y/O TIEMPO

C CUESTA, I BARBONA, G ISERN y N MEROI

FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA, UNR.

ccuesta@fcecon.unr.edu.ar

La denominada “falacia ecológica” en estudios cuya unidad de observación es un área geográfica (o tiempo), y donde la variable observada es una estadística resumen (promedio, porcentaje, tasa, etc), se refiere al sesgo que se comete al querer extrapolar las asociaciones observadas a nivel de área, a los individuos de las mismas. Este tipo de sesgos puede producirse debido a diferentes motivos: la incapacidad de determinar la temporalidad de la variable respuesta y la explicativa, la posible presencia de variables confundentes, la imposibilidad de contar con las distribuciones de las medidas agrupadas, etc. A pesar del conocimiento de estos problemas, los estudios ecológicos siguen siendo muy utilizados en diversas áreas debido a su sencillez, bajo costo, fácil obtención de la información, etc. y son utilizados, generalmente en una faz exploratoria. A fin de controlar el posible sesgo asociado a estos estudios se han propuesto diferentes modelos, entre ellos los que tienen en cuenta estratos definidos de acuerdo a alguna posible variable de confusión. La ventaja de estos modelos estratificados es que se trabaja con grupos más pequeños que el total del área geográfica permitiendo tener en cuenta posibles variables de confusión (que son las que determinan los estratos). Sin embargo, esta metodología puede no ser adecuada cuando se carece de la información a nivel de estrato y/o cuando las áreas de estudio son demasiado pequeñas. En este trabajo se presenta un modelo que pretende controlar el mencionado sesgo combinando distintas fuentes de datos, una que incluye los datos agrupados y otra que incluye los datos a nivel individual (que puede no ser la misma que la anterior). Este modelo es un intermedio entre utilizar datos a nivel individuo y datos a nivel área geográfica. Se comparan los resultados de este modelo con los que resultan bajo un modelo estratificado y con uno que considera los individuos como unidad de análisis. Se ejemplifica la metodología utilizando datos de la Encuesta Nacional de Factores de Riesgo llevada a cabo por las Direcciones Provinciales de Estadística entre octubre y diciembre de 2013. Se estudia la asociación entre obesidad y diferentes factores tales como edad, sexo, nivel de instrucción, cobertura de salud, etc.

Palabras claves: falacia ecológica; estudios ecológicos; epidemiología

RE-PENSANDO UN ESTUDIO CONFIRMATORIO COMO “ADAPTATIVO” USANDO EL ENFOQUE DE LA PROBABILIDAD PREDICTIVA Y LOS MODELOS DE REGRESSION POR CUANTILES

CARMEN VALENZUELA¹, PATRICIA LORENZO-LUACES¹, CAMILO RODRÍGUEZ¹, ELIA NENINGER², ARASAY MONTES¹ y TANIA CROMBET¹

¹ CENTRO DE INMUNOLOGÍA MOLECULAR (CIM), CUBA.. ²HOSPITAL CLÍNICO QUIRÚRGICO HERMANOS AMEJERIAS, CUBA.

carmenv@cim.sld.cu

Introducción: Los Ensayos Clínicos Adaptativos son aceptados y reconocidos actualmente. Sin embargo, algunas veces las agencias reguladoras no son entusiastas con las decisiones basadas en estrategias de simulación. El enfoque de la Probabilidad Predictiva ofrece un modo natural de estimar el éxito en el análisis futuro, condicionado a los datos acumulados. Por otro lado, la regresión

por cuantiles es de gran flexibilidad en la evaluación del efecto de covariables sobre una variable “tiempo hasta la ocurrencia de un evento”, como análisis complementario a los modelos de regresión de Cox. Desarrollo: Los datos acumulados de 6 estudios exploratorios fueron considerados como información previa para re-pensar el estudio fase III como adaptativo, estimando iterativamente la probabilidad predictiva de éxito basados en las distribución final (análisis bayesiano) y usando el enfoque frecuentista. Se compararon los resultados del modelo de regresión de Cox tradicional con los resultados de la regresión por cuantiles. Principales resultados: Usando ambos enfoques para estimar la probabilidad predictiva, la probabilidad de éxito fue mayor de 0.95 con el 25% del tamaño de muestra planificado para el estudio fase III, consistente con los resultados finales. El análisis bayesiano muestra algunas ventajas respecto al enfoque frecuentista, considerando diferentes maneras de usar la información previa. Por otro lado, la regresión por cuantiles fue de gran utilidad para comprender el efecto de las covariables en el tiempo de supervivencia a los diferentes niveles de los cuantiles. La influencia de algunas covariables no se mantiene a lo largo de toda la distribución. Conclusiones: Estas herramientas pueden ser cruciales en el proceso de toma de decisiones en enfermedades raras, cuando las alternativas de tratamiento son no efectivas o no accesibles, o cuando existe hetetocedasticidad en los datos. El enfoque de la probabilidad predictiva puede optimizar el diseño de futuros ensayos y contribuir a una mayor eficiencia de los mismos. El supuesto de riesgos proporcionales constituye una restricción al uso de los modelos de Cox y en ese sentido la regresión por cuantiles es más flexible, natural e informativa, permitiendo identificar y modelar las causas de heterogeneidad que impactan en la variable dependiente. Introducción: Los Ensayos Clínicos Adaptativos son aceptados y reconocidos actualmente. Sin embargo, algunas veces las agencias reguladoras no son entusiastas con las decisiones basadas en estrategias de simulación. El enfoque de la Probabilidad Predictiva ofrece un modo natural de estimar el éxito en el análisis futuro, condicionado a los datos acumulados. Por otro lado, la regresión por cuantiles es de gran flexibilidad en la evaluación del efecto de covariables sobre una variable “tiempo hasta la ocurrencia de un evento”, como análisis complementario a los modelos de regresión de Cox. Desarrollo: Los datos acumulados de 6 estudios exploratorios fueron considerados como información previa para re-pensar el estudio fase III como adaptativo, estimando iterativamente la probabilidad predictiva de éxito basados en las distribución final (análisis bayesiano) y usando el enfoque frecuentista. Se compararon los resultados del modelo de regresión de Cox tradicional con los resultados de la regresión por cuantiles. Principales resultados: Usando ambos enfoques para estimar la probabilidad predictiva, la probabilidad de éxito fue mayor de 0.95 con el 25% del tamaño de muestra planificado para el estudio fase III, consistente con los resultados finales. El análisis bayesiano muestra algunas ventajas respecto al enfoque frecuentista, considerando diferentes maneras de usar la información previa. Por otro lado, la regresión por cuantiles fue de gran utilidad para comprender el efecto de las covariables en el tiempo de supervivencia a los diferentes niveles de los cuantiles. La influencia de algunas covariables no se mantiene a lo largo de toda la distribución. Conclusiones: Estas herramientas pueden ser cruciales en el proceso de toma de decisiones en enfermedades raras, cuando las alternativas de tratamiento son no efectivas o no accesibles, o cuando existe hetetocedasticidad en los datos. El enfoque de la probabilidad predictiva puede optimizar el diseño de futuros ensayos y contribuir a una mayor eficiencia de los mismos. El supuesto de riesgos proporcionales constituye una restricción al uso de los modelos de Cox y en ese sentido la regresión por cuantiles es más flexible, natural e informativa, permitiendo identificar y modelar las causas de heterogeneidad que impactan en la variable dependiente.

Palabras claves: diseños adaptativos; probabilidad predictiva; regresión por cuantiles

UN MODELO DE REGRESIÓN PARA EL ANÁLISIS DE BIOMASA DE ALGAS MARINAS

JOSÉ SARAVIA

SECRETARÍA DE PESCA DE LA PROVINCIA DEL CHUBUT.

josesaravia@speedy.com.ar

Este trabajo presenta una propuesta de aplicación de regresión heteroscedástica para evaluar la biomasa de algas bentónicas marinas y su relación con diversas variables ambientales y nutrientes. Se evaluaron tres praderas existentes en la zona norte del golfo San Jorge, con diferentes cantidades de puntos de muestreo (estaciones). La biomasa se midió a lo largo de 9 campañas, registrándose conjuntamente las variables pH, salinidad, oxígeno disuelto, temperatura, nitratos y fosfatos del agua de mar. Se realizó una regresión común y ante la desviación de los supuestos del modelo (homoscedacia y normalidad), se optó por aplicar modelos lineales mixtos con la variable clasificatoria "estaciones" como aleatoriedad. El gráfico correspondiente muestra que la distribución de los residuos es aproximadamente normal. Al analizar el diagrama de dispersión de los residuos condicionales de Pearson versus los Valores ajustados, se observa una clara tendencia de los residuos a aumentar su varianza a medida que aumenta el valor medio. Esto sugiere la necesidad de modelar esta falta de homogeneidad de varianza con una función que relacione las varianzas de los residuos con la media. Se plantea el mismo modelo pero con diversas funciones de varianzas. Se concluye que la regresión puede ser utilizada para describir el comportamiento de la biomasa de algas ya que el ajuste fue satisfactorio considerando que la varianza es distinta para cada estación. La biomasa de algas marinas está relacionada con pH, oxígeno disuelto y temperatura. No se observaron diferencias significativas entre praderas. La mayor variabilidad es atribuible a diferencias entre estaciones.

Palabras claves: regresión heterosced; biomasa; modelo lineal mixto

Muestreo

DIMENSIONES CONCEPTUALES Y METODOLÓGICAS DEL ESTUDIO POR MUESTREO A PRODUCTORES HORTÍCOLAS EN LOS DEPARTAMENTOS DE LUJÁN Y EXALTACIÓN DE LA CRUZ, PROVINCIA DE BUENOS AIRES

HUGO DELFINO¹, VIVIANA ESCANES² y OLGA SUSANA FILIPPINI¹

¹UNLU - CS. BÁSICAS. ²UNLU - CS. SOCIALES.

h.delfino@yahoo.com.ar

La producción hortícola argentina se caracteriza por su amplia distribución geográfica y por la diversidad de especies producidas. Analizar la implementación de "buenas prácticas agropecuaria y las características de la producción hortícola del partido de Luján y Exaltación de la Cruz resulta de suma utilidad para los organismos que difunden Tecnologías Agropecuarias y los municipios del periurbano bonaerense. En torno a los grandes núcleos urbanos se han desarrollado los llamados "cinturones verdes": territorios ocupados por quintas o huertas familiares y comerciales que

rodean a las ciudades y donde se producen hortalizas para abastecer a la población urbana. Estos cinturones se caracterizan por su tamaño relativamente pequeño o mediano de las explotaciones (1 a 40 hectáreas) y por la amplia variedad de especies cultivadas. La expansión urbana expulsó a las “familias productoras” a regiones más alejadas y precarias, con escasa planificación territorial y menor acceso a servicios básicos. Esta nueva población periurbana también incluye a las familias de productores y trabajadores rurales provenientes de migraciones internas y externa, principalmente del Norte argentino y de Bolivia. Describir la producción hortícola en relación con el uso del suelo, tipo de producción y régimen de tenencia, conocer las estrategias y formas de inserción en el área productiva y relevar la utilización de Buenas Prácticas aplicables a la producción y comercialización por parte de los actores claves, fue el objetivo de este estudio. Se realiza una encuesta por muestreo, con selección al azar de productores a partir de un listado proporcionado por el Instituto Nacional de Tecnología Agropecuaria. El relevamiento de la información permite identificar aspectos socio-demográficos, nivel de Buenas Prácticas de producción Agrícolas, origen y destino de la producción, cambios posibles en la dinámica productiva de la zona bajo estudio que sirve de insumo para investigadores de la Universidad Nacional de Luján y el Instituto Nacional de tecnología Agropecuario. Los resultados que se muestran tienen carácter de provisarios pero la representatividad de la muestra se está consolidando, ya que “en grupos bien seleccionados, con percepciones e intereses comunes sobre un tema, con un tamaño adecuado de entrevistas, se pueden detectar patrones comunes de la población bajo estudio, ya que la curva de aparición de nuevos conceptos alcanza una meseta.

Palabras claves: muestreo probabilístico; periurbano; productores hortícol

ESTIMACIÓN POR MUESTREO DE LAS SUPERFICIES SEMBRADAS, CON CULTIVOS EXTENSIVOS MÉTODO DE SEGMENTOS ALEATORIOS

CECILIA CONDA

DIRECCIÓN DE INFORMACIÓN AGROPECUARIA Y FORESTAL - SUBSECRETARIA DE AGRICULTURA, MINISTERIO DE AGROINDUSTRIA..

ceciliaconde@gmail.com

El conocimiento de las superficies sembradas con cultivos de tipo extensivo es de relevancia estratégica para el país y se hace necesario estimarlas para todas aquellas zonas y regiones de las Provincias Argentinas que revisten importancia agropecuaria. Asimismo es la base para la determinación de la producción agrícola que es requerida por numerosos actores económicos, sociales y políticos para planificar sus acciones, reducir incertidumbre y mejorar la asignación de los recursos. Las estimaciones agropecuarias deben satisfacer una serie de requisitos para poder asegurar su confiabilidad y estar disponible en tiempo y forma es decir, en el momento en que se necesite para tomar decisiones. Como consecuencia el Ministerio de Agroindustria, desde la Dirección de Información Agropecuaria y Forestal, desarrolló e implementó desde la campaña agrícola 2011/2012 el Método de Segmentos Aleatorios (MSA). El MSA consiste en un diseño de muestra estratificado de conglomerados monoetápico dentro de cada departamento o partido del país. La unidad de muestreo es el segmento, el cual consiste en un área de 400has. con diferentes coberturas del suelo, ubicado aleatoriamente en estratos homogéneos según la intensidad de uso agrícola, la cual es analizada a través de información satelital. Los operativos de recolección de datos se realizan de forma bianual coincidente con las campañas agrícolas de cosecha fina y de cosecha gruesa. El relevamiento consiste en observaciones visuales directas de los usos del suelo sin consultar a productores o informantes de las explotaciones, minimizando así los errores no debidos al muestreo. Se definieron a priori más de 30 diferentes tipos de cobertura. Esta información es integrada en un sistema de información geográfica (SIG), el cual hace posible calcular exactamente la superficie de cada tipo de cobertura dentro del segmento. Los estimadores utilizados para expandir la

superficie de las coberturas muestrales son en su mayoría estimadores basados en la metodología de Horvitz-Thompson. Circunstancialmente cuando los segmentos son de tamaño variable se usan estimadores por razón a la superficie del segmento. Los resultados son informados a través de estimadores puntuales y por intervalos de confianza, adicionando sus correspondientes errores estándar y los coeficientes de variación. El conocimiento de las superficies sembradas con cultivos de tipo extensivo es de relevancia estratégica para el país y se hace necesario estimarlas para todas aquellas zonas y regiones de las Provincias Argentinas que revisten importancia agropecuaria. Asimismo es la base para la determinación de la producción agrícola que es requerida por numerosos actores económicos, sociales y políticos para planificar sus acciones, reducir incertidumbre y mejorar la asignación de los recursos. Las estimaciones agropecuarias deben satisfacer una serie de requisitos para poder asegurar su confiabilidad y estar disponible en tiempo y forma es decir, en el momento en que se necesite para tomar decisiones. Como consecuencia el Ministerio de Agroindustria, desde la Dirección de Información Agropecuaria y Forestal, desarrolló e implementó desde la campaña agrícola 2011/2012 el Método de Segmentos Aleatorios (MSA). El MSA consiste en un diseño de muestra estratificado de conglomerados monoetápico dentro de cada departamento o partido del país. La unidad de muestreo es el segmento, el cual consiste en un área de 400has. con diferentes coberturas del suelo, ubicado aleatoriamente en estratos homogéneos según la intensidad de uso agrícola, la cual es analizada a través de información satelital. Los operativos de recolección de datos se realizan de forma bianual coincidente con las campañas agrícolas de cosecha fina y de cosecha gruesa. El relevamiento consiste en observaciones visuales directas de los usos del suelo sin consultar a productores o informantes de las explotaciones, minimizando así los errores no debidos al muestreo. Se definieron a priori más de 30 diferentes tipos de cobertura. Esta información es integrada en un sistema de información geográfica (SIG), el cual hace posible calcular exactamente la superficie de cada tipo de cobertura dentro del segmento. Los estimadores utilizados para expandir la superficie de las coberturas muestrales son en su mayoría estimadores basados en la metodología de Horvitz-Thompson. Circunstancialmente cuando los segmentos son de tamaño variable se usan estimadores por razón a la superficie del segmento. Los resultados son informados a través de estimadores puntuales y por intervalos de confianza, adicionando sus correspondientes errores estándar y los coeficientes de variación. El conocimiento de las superficies sembradas con cultivos de tipo extensivo es de relevancia estratégica para el país y se hace necesario estimarlas para todas aquellas zonas y regiones de las Provincias Argentinas que revisten importancia agropecuaria. Asimismo es la base para la determinación de la producción agrícola que es requerida por numerosos actores económicos, sociales y políticos para planificar sus acciones, reducir incertidumbre y mejorar la asignación de los recursos. Las estimaciones agropecuarias deben satisfacer una serie de requisitos para poder asegurar su confiabilidad y estar disponible en tiempo y forma es decir, en el momento en que se necesite para tomar decisiones. Como consecuencia el Ministerio de Agroindustria, desde la Dirección de Información Agropecuaria y Forestal, desarrolló e implementó desde la campaña agrícola 2011/2012 el Método de Segmentos Aleatorios (MSA). El MSA consiste en un diseño de muestra estratificado de conglomerados monoetápico dentro de cada departamento o partido del país. La unidad de muestreo es el segmento, el cual consiste en un área de 400has. con diferentes coberturas del suelo, ubicado aleatoriamente en estratos homogéneos según la intensidad de uso agrícola, la cual es analizada a través de información satelital. Los operativos de recolección de datos se realizan de forma bianual coincidente con las campañas agrícolas de cosecha fina y de cosecha gruesa. El relevamiento consiste en observaciones visuales directas de los usos del suelo sin consultar a productores o informantes de las explotaciones, minimizando así los errores no debidos al muestreo. Se definieron a priori más de 30 diferentes tipos de cobertura. Esta información es integrada en un sistema de información geográfica (SIG), el cual hace posible calcular exactamente la superficie de cada tipo de cobertura dentro del segmento. Los estimadores utilizados para expandir la superficie de las coberturas muestrales son en su mayoría estimadores basados en la metodología de Horvitz-Thompson. Circunstancialmente cuando los segmentos son de tamaño variable se usan estimadores por razón a la superficie del segmento. Los resultados son informados a través de estimadores puntuales y por intervalos de confianza, adicionando sus correspondientes errores estándar y los coeficientes de variación.

Palabras claves: superficie agrícola sembrada; segmentos aleatorios; uso del suelo; estimadores por simple expansión y razón

Probabilidad y procesos estocásticos

DEMORA EN LA ELECCIÓN DE CARRERA EN LA FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA UNA ESTIMACIÓN MEDIANTE CADENAS DE MARKOV

GABRIELA BOGGIO, NORA ARNESI y LUCIANA RUIZ

INSTITUTO DE INVESTIGACIONES TEÓRICAS Y APLICADAS DE LA ESCUELA DE ESTADÍSTICA, FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA, UNIVERSIDAD NACIONAL DE ROSARIO.

gboggio@fcecon.unr.edu.ar

Uno de los problemas que enfrenta la Facultad de Ciencias Económicas y Estadística de la Universidad Nacional de Rosario en relación al rendimiento académico de los alumnos es la demora en los tiempos previstos por los planes de estudio para cumplir una primera meta relacionada con la posibilidad de elegir carrera. El plan de estudios de las tres carreras de Ciencias Económicas de esta Facultad comienza con un Ciclo Introductorio Común y la elección de la carrera es posible previa aprobación de cuatro asignaturas de este ciclo. Por ello interesaría estimar el tiempo promedio que los alumnos demoran hasta alcanzar esta condición requerida para elegir carrera. A tal fin se analizó la cohorte de ingresantes del año 2008, evaluando al inicio de cada año académico la cantidad de materias aprobadas de ese ciclo introductorio hasta la finalización del año académico 2016. Se tomó como criterio que el alumno que estuvo pasivo durante dos años académicos consecutivos está en una situación no explícita asimilable al abandono de la carrera. Se recurrió a la metodología de cadenas de Markov definiendo como estados absorbentes los que representan estar en condición de elegir carrera o abandonar los estudios y como estados transitorios la cantidad de materias aprobadas al finalizar cada año académico (ninguna, una, dos o tres). Se pudo evidenciar la falta de homogeneidad en las probabilidades de transición entre los distintos estados en los 9 años siguientes al ingreso de los alumnos, fundamentalmente en el primer año de cursado. Para salvar este problema se usaron los denominados estados túneles que permiten establecer probabilidades de transición con un ajuste temporal y se pusieron a prueba distintas estructuras de cadena teniendo en cuenta reagrupamientos de los estados en la búsqueda de la mejor representación de la situación estudiada. El tiempo promedio hasta la elección de carrera bajo las diferentes estructuras consideradas resultó igual a 2,3 años y las diferentes representaciones dieron lugar a la estimación de la probabilidad de alcanzar esta condición bajo distintas circunstancias. Por ejemplo, la probabilidad de que un ingresante logre la condición en el primer año resultó igual a 0,57 mientras que aquél que, finalizado el primer año, no pudo aprobar materias tiene una probabilidad de 0,08 de lograr este objetivo. Dicha probabilidad se redujo a 0,02 ante la persistencia en esta situación pasados dos ciclos.

Palabras claves: cadenas absorbentes; estados túneles; rendimiento académico

LA ENTROPÍA COMO MÉTODO PARA COMPARAR IMÁGENES EN EL TIEMPO

SILVIA P MELO¹ y RODOLFO C J SALOMÓN²

¹DEPARTAMENTO DE MATEMÁTICA, UNIVERSIDAD NACIONAL DEL SUR, ARGENTINA. ²DEPARTAMENTO DE GEOLOGÍA, UNIVERSIDAD NACIONAL DEL SUR, ARGENTINA.

smelo@uns.edu.ar

Una sucesión de imágenes, que ofrecen información respecto a un determinado proceso, pueden ser abordadas desde diversos puntos de vista. Al analizar los cambios que se producen a lo largo del tiempo, resulta de interés poder determinar en qué momento comienzan a aparecer diferencias que puedan ser consideradas significativas. En esta instancia, es deseable poder obtener resultados que no dependan de la experiencia y / o criterio del observador. En el presente trabajo se busca establecer un método automático y repetible en el tiempo, accesible en su interpretación, de fácil implementación y aplicación, tanto para el usuario entrenado como para el no entrenado. El principio es la aplicación del concepto de entropía . Se considera la información subyacente en cada imagen, como una matriz de m x n puntos de resolución, cuyos elementos corresponden a los valores de luminosidad tomados en escala de grises. En base a esta matriz, se estima la distribución de probabilidades de la escala de grises por medio de las frecuencias relativas observadas, lo que permite el cálculo de la entropía. El método propuesto es: tomando como punto de partida la entropía de la imagen inicial, considerar las diferencias con los correspondientes valores de entropía de las restantes imágenes. El criterio para tal comparación se fundamenta en que pequeños cambios generados por el tratamiento aplicado, no modificarán sustancialmente la distribución de probabilidades inicial. Cuando estos cambios sean mayores, la disociación entre las imágenes quedará evidenciada, tornándose significativa la diferencia entre las entropías. Se presentan resultados obtenidos mediante simulación y por aplicación a datos reales

Palabras claves: entropía; asociación; imagen; probabilidad

Probabilidades

PREDICCIÓN DE LA DISTRIBUCIÓN DEL TAMAÑO DE FRUTOS DE TANGOR 'MURCOTT' A COSECHA

G.R.R. BÓBEDA¹, L.I. GIMÉNEZ¹, SM MAZZA¹ y SJ BRAMARDI²

¹CÁTEDRA DE CÁLCULO ESTADÍSTICO Y BIOMETRÍA. FACULTAD DE CIENCIAS AGRARIAS. UNIVERSIDAD NACIONAL DEL NORDESTE. CORRIENTES, ARGENTINA. ²DEPARTAMENTO DE ESTADÍSTICA. UNIVERSIDAD NACIONAL DE COMAHUE. NEUQUÉN, ARGENTINA. CITAAC (CENTRO DE INVESTIGACIONES EN TOXICOLOGÍA AMBIENTAL Y AGROBIOTECNOLOGÍA DEL COMAHUE). CONICET-UNCO.

griseldabobeda@gmail.com

La estimación temprana del volumen de producción y distribución de tamaños de frutos en cada temporada es una herramienta fundamental en la definición de prácticas de manejo de cultivos, planificación de cosecha y destinos de comercialización de frutos cítricos. Existen diferentes formas de realizar dicha estimación., una de ellas es analizar la distribución de probabilidad que presenta el tamaño con dos meses de anticipación y utilizar el ajuste de dicha distribución para el cálculo de probabilidades por calibre comercial. El presente trabajo se realizó con el objetivo de analizar

la distribución que adoptan los calibres comerciales en 8 lotes de tangor 'Murcott' en las localidades de Bella Vista, Concepción y San Lorenzo, durante 5 campañas en la provincia de Corrientes, Argentina. La información se obtuvo de las mediciones del diámetro equatorial (mm) de frutos de mandarino tangor Murcott, al momento de la cosecha a partir de una muestra de 50 frutos de 10 árboles por lote comercial. Los datos se agruparon según las categorías de calibres comerciales utilizadas para el mercado externo según el boletín oficial de la Unión Europea, que varían de 53 a 93.5 mm y el número de frutos promedios oscila entre 48 a 130 frutos para cajas plató de 10 kg. Luego de la categorización, se analizó la variabilidad de las distribuciones obtenidas para cada lote y campaña. Mediante herramientas de Estadísticas Descriptiva y para evaluar el ajuste de la distribución de tamaños a una distribución normal se utilizó la Prueba de Bondad de ajuste de χ^2 (Chi-Cuadrado). Luego se procedió a construir tablas de predicción de porcentajes de la cantidad de frutos en función de la media y desvío estándar del diámetro equatorial. Las tablas construidas contienen medias de diámetros que van desde 53 a 82 mm, con desvíos que varían de 4 a 6 mm. que podrían estar asociados a diferentes condiciones ambientales y de manejo. Este ajuste a la distribución normal del diámetro equatorial permitió la construcción de tablas de predicción de porcentajes de frutas de cada calibre comercial. Para estimar el porcentaje de calibres de calibres comerciales, se medirá los diámetros con 2 meses de anticipación en cada campaña, y posteriormente predecir la distribución de los tamaños a cosecha. Esto constituye una herramienta útil para productores y sus futuras planificaciones de manejos culturales y comercialización de la fruta.

Palabras claves: prueba de bondad de ajuste; calibres comerciales; curvas de crecimiento

Series de tiempo

ESTUDIO DE LA DINÁMICA DEL BENEFICIO ENERGÉTICO BRUTO ANUAL EN SISTEMAS DE PASTOREO TRASHUMANTE

NATALIA PEREZ, MARCOS EASDALE y OCTAVIO BRUZZONE

CONICET - INTA.

nperezleon@gmail.com

El objetivo de este trabajo fue desarrollar un método para estimar el beneficio energético bruto anual (BEBa) para la producción ganadera trashumante, al moverse entre sitios de invernada y veranada. En el pastoreo trashumante las personas y sus rebaños se desplazan estacionalmente desde las tierras bajas, que son pastoreadas durante el invierno (invernadas), hacia las tierras altas situadas en zonas montañosas durante el verano para aprovechar pastos de alta calidad (veranadas). Se seleccionaron cuatro casos biofísicamente contrastantes entre sitios de pastoreo trashumante de la Patagonia noroccidental, Argentina, luego se delimitaron las invernadas y veranadas. Series de imágenes MODIS compuestas cada 16 días, desde febrero de 2000 a septiembre de 2016 (píxeles de 250m x 250m) conteniendo el Índice de Vegetación Diferencial Normalizado (NDVI), fueron usadas como estimador indirecto de la productividad primaria neta aérea (Rouse et al., 1974). Luego estimamos el BEBa como las diferencias absolutas entre las curvas anuales de NDVI de cada veranada e invernada asociadas a un productor trashumante, respectivamente. Luego, el BEB parcial (BEBp) de pasar de una invernada a una veranada (y viceversa), es la diferencia en la productividad estacional calculada cuando un lugar es más productivo que el otro. Se hizo un filtrado de paso de múltiples bandas para obtener el ciclo anual y sus armónicos, así como un filtrado de paso bajo (frecuencias menores a un año), para obtener la variabilidad interanual de la estimación de BEB. Sobre las series

filtradas, se estimó el BEBa, para cada año. Entre los resultados más relevantes obtuvimos que el BEBa fue siempre positivo en la mayoría de los casos estudiados, sugiriendo que la movilidad del ganado entre sitios de invernada y veranada es beneficiosa. La variabilidad entre valores sucesivos del BEBa fue baja (menor al 5%) y estable en el tiempo, corroborando que el grado de incertidumbre para los productores sea relativamente pequeño. Este método permite una estimación rápida del BEBa y su variabilidad en el tiempo (a lo largo de los años). El mismo podría servir de base para desarrollar un sistema de seguimiento y alerta temprana de cambios en el BEB en un momento determinado, con el fin de anticipar decisiones productivas.

Palabras claves: conectividad; fragmentación; ndvi; socio-ecológicos; transformada de four; variabilidad

SELECCIÓN DE ESTRUCTURAS DE (CO)VARIANZAS EN EL ANÁLISIS DE DATOS EN ENSAYOS SILVICULTURALES

CESAR GASTÓN TORRES¹, ESTEBAN JIMÉNEZ ALFARO² y DANIELA
LAURA PICARDI³

¹EEA INTA BELLA VISTA. ²ESCUELA DE CIENCIAS AGRARIAS, UNIVERSIDAD NACIONAL COSTA RICA. ³DEPARTAMENTO DE PRODUCCIÓN ANIMAL, FACULTAD DE AGRONOMÍA - UBA.

torresgastonc@gmail.com

El estudio de la producción forestal requiere de estimaciones de crecimiento que brinden información de los recursos actuales, además dichas estimaciones deben predecir de manera fiable el comportamiento futuro de los mismos. Para ello, se establecen ensayos que funcionan durante períodos preestablecidos y generan información en el tiempo de variables dasométricas de interés, como es el caso de la Diámetro Normal (Dn). Durante el ajuste se utilizan múltiples mediciones en una misma unidad de observación, ello determina la falta de independencia entre mediciones provenientes de la misma unidad, las cuales están autocorrelacionadas con un grado determinado por el método de medición. Mediante técnicas que maximizan la verosimilitud del modelo es posible modelar la autocorrelación implícita o explícitamente. En el presente trabajo, se llevó a cabo el ajuste de modelos mixtos de datos longitudinales con diferentes estructuras de varianzas y covarianzas, para el Dn. Se analizó datos de un ensayo de preparación de suelo (4 tipos) con tres materiales comerciales de pino, en un diseño Split Plot con cuatro bloques completamente aleatorizados. Sobre la parcela principal se asignó a la preparación del terreno y los materiales a la subparcela. La variable analizada se midió a las edades de 34, 49, 74, 93 y 114 meses. Se evaluó un modelo mixto con efectos fijos (edad, tratamiento y materiales) con sus respectivas interacciones y efectos aleatorios de bloque (b) y parcela anidada en bloque (u). En el modelo alternativo se eliminó las interacciones no significativas de los efectos fijos. Ambos modelos se ajustaron considerando homocedasticidad y varianzas homogéneas dentro de edad. Las estructuras de correlación evaluadas corresponden a: autorregresiva de primer orden (AR1), autorregresiva continua de orden uno (CAR1) y totalmente desestructurada (UN). La selección del modelo se realizó mediante el Criterio de Información de Akaike (AIC). El modelo que mostró el menor valor de AIC (665.5) fue la variante reducida con componentes de varianza homogéneas dentro de edad, al ajustar la autocorrelación de forma autorregresiva a nivel de subparcela. Los parámetros estimados fueron $u_2=0.41$ (varianza de u), $e_2=2.85$ (varianza residual) y $= 0.558$ (coeficiente de correlación). Los coeficientes de ponderación de los componentes de la varianza fueron 1, 0.39, 0.39, 0.54 y 0.58 para las cinco edades respectivas. Los resultados muestran que la inclusión de la matriz de (co)varianzas disminuye 99 puntos el valor de AIC.

Palabras claves: medidas repetidas; máxima verosimilitud; modelos lineales mix; observaciones no equ

Otras categorías metodológicas

ANÁLISIS DESCRIPTIVO LONGITUDINAL DE LA FRECUENCIA DE INTIMIDACIÓN ENTRE PARES Y SUS PRINCIPALES CARACTERÍSTICAS EN ESCOLARES DE BAHÍA BLANCA

MARÍA GABRIELA SERRALUNGA, CECILIA VASCONI y LUCAS DURÁN

UNIVERSIDAD NACIONAL DEL SUR.

mg.serra@criba.edu.ar

INTRODUCCIÓN: El objetivo del trabajo es estimar longitudinalmente la frecuencia de la intimidación entre pares, según la percepción de los alumnos, y analizar sus principales características.

METODOLOGÍA: Estudio longitudinal descriptivo realizado en niños de 8 a 12 años, de siete escuelas primarias de gestión estatal y privada de la ciudad de Bahía Blanca, seleccionadas mediante un muestreo por conveniencia, durante mayo (Toma 1) y noviembre (Toma 2) de 2016. Instrumento de recolección de datos: Cuestionario auto-administrado “Pre-Concepciones de Intimidación y Maltrato entre Iguales” (PRECONCIMEI, Avilés, 2002). Definición de roles según la participación del niño en situaciones de intimidación, recibiendo maltrato, ejerciéndolo, o ambos: Participación sostenida si manifestó haber participado en ambas tomas, Participación episódica si indicó participar en sólo una, Sin participación activa si manifestó no haber participado en ninguna toma. Las características de la intimidación se analizaron sólo en la primera toma. Programa utilizado SPSS 17. **RESULTADOS:** De los 458 niños/as encuestados, un 30,3% y un 35,2% refirieron participar en situaciones de intimidación en mayo y noviembre, respectivamente. El análisis longitudinal reveló que un 17,9% participó sostenidamente, un 29,6% participó episódicamente y el 52,4% restante no participó activamente en ninguna de las tomas. Se encontraron diferencias significativas al comparar estos porcentajes a través del tiempo (años 2012, 2014, 2015 y 2016). En cuanto a sus características, el tipo de intimidación más frecuente en niños fue “Insultar y poner apodos” (58,2%) y en niñas “Hablar mal de alguien” (48,1%); el lugar donde ocurren con mayor frecuencia fue “En el patio cuando no vigilaba ninguna maestra” (14,4%); para ambos sexos el interlocutor válido a quien recurren fue “la familia” (28,2%); quien más frecuentemente suele parar estas situaciones fue “alguna maestra” (42%); respecto a porqué lo hacen, la mayoría (53,2%) dijo “por molestar”; qué hicieron cuando observaron tales situaciones, “Nada, no me metí” (26,2%), y la respuesta más frecuente a qué habría que hacer para solucionar el problema, fue “Que hagan algo las maestras” (34,3%). **CONCLUSIÓN:** La observación en dos momentos del ciclo escolar permitió detectar que la mayoría de los niños no participa de estas situaciones en forma activa en ningún momento del año, un 30% cambia de perfil, entrando o saliendo de tales situaciones con un patrón dinámico, y un 18% participa en forma sostenida. El conocimiento de esta dinámica es de utilidad tanto para el diseño de intervenciones escolares, como para el abordaje preventivo de la problemática.

Palabras claves: intimidación entre p; encuesta preconcimei; escolares

ANALISIS Y EVALUACIÓN DE LA PRODUCCIÓN DE CAÑA DE AZUCAR MEDIANTE SIMEX

NELIDA DEL VALLE ORTIZ, PATRICIA ANDREA DIGONZELLI,, MARIA BEATRIZ GARCIA, JUAN FERNANDEZ DE ULLIVARRI y EDUARDO ROMERO

UNIVERSIDAD NACIONAL DE TUCUMÁN.

ortiznelidadelvalle@gmail.com.ar

La caña de azúcar es un cultivo perenne cuya mejor producción es durante los cinco primeros años del cultivo. Concluida la cosecha los residuos agrícolas pueden dejarse en la superficie o eliminarse. Para comparar las producciones de ambas prácticas agrícolas se registró el número de tallos a lo largo del ciclo del cultivo. Por tratarse de datos longitudinales de conteo el análisis se realizó mediante ecuaciones de estimación generalizadas con estructura de varianza intercambiable. El objetivo de este trabajo fue analizar el error de medida en el número tallos y evitar estimaciones sesgadas de los parámetros. Las observaciones se realizaron durante el ciclo del cultivo en días posteriores a la cosecha en diferentes frecuencias de tiempo generando desiguales intervalos aleatorios. El análisis descriptivo de los datos mostró que la el número de tallos en parcelas sin residuos agrícolas es superior hasta, aproximadamente, la mitad del ciclo y decrece manteniéndose entre valores próximos hasta el final del ciclo. Simultáneamente, la producción correspondiente a parcelas con residuos agrícolas permanece con valores inferiores y también decrece hacia el final. Por lo tanto, se decidió analizar los datos desde las últimas instancias del cultivo marcado por el comienzo del decrecimiento de la producción del número de tallos en parcelas sin cobertura agrícola. Se determinó como variables de predicción al tiempo transcurrido desde la cosecha, es decir, a días posteriores a la cosecha y la ausencia o presencia de la cobertura de residuos agrícola como el tratamiento aplicado. Naturalmente, la variable respuesta fue el número de tallos. Los resultados mostraron homogeneidad entre las producciones de ambas prácticas agrícolas. Para el análisis de error en la variable dependiente se aplicó el método de inferencia para modelos con error de medida basado en simulación y extrapolación SIMEX, que por tratarse de datos de conteo se usó MCSIMEX. Se compararon las desviaciones estándar de las estimaciones de los parámetros del modelo de ecuaciones de estimación generalizadas en el que se asume que no existe error de medida, es decir modelo naive, y el modelo mcsimex generado desde la aplicación del método MCSIMEX en el modelo naive. De la comparación se observaron mínimas diferencias entre las desviaciones estándar.

Palabras claves: datos longitudinales; ecuaciones de estima; simex; variable discreta

APLICACIÓN DE ANALISIS FACTORIAL MÚLTIPLE SOBRE DATOS DE CALIDAD DE AGUA MEDIDA MEDIANTE DIFERENTES GRUPOS DE INDICADORES

L MAYANS¹, S OCAMPO¹, J MOLINA¹, L LAFFITTE², J MUÑIZ SAAVEDRA² y A LAVALLE¹

¹UNIVERSIDAD NACIONAL DEL COMAHUE. ²SUBSECRETARÍA DE AMBIENTE.

lau.mayans@hotmail.com

La cuenca Lacar – Hua Hum, constituye la única cuenca de vertiente pacífica de la Provincia del Neuquén. Durante el año 2016 se llevó a cabo el monitoreo bio-hidrológico de la misma, en colaboración entre la Dirección General de Biología Acuática, Subsecretaría de Ambiente de Neuquén,

quién financió el monitoreo y el Organismo de Control Municipal de San Martín de los Andes. Se muestrearon 29 sitios de diferentes tipos de cuerpos y cursos de agua (arroyos, ríos, lago, canales) entre los meses de marzo y abril de 2016. Se determinaron parámetros físico-químicos, biológicos y de hábitat de cada sitio. Los resultados se utilizaron para determinar calidad de sitios en base a bioindicadores y para comparar la evolución de la calidad en base a datos históricos de algunos de los sitios. En el presente trabajo se utilizaron datos de 21 de los sitios mencionados. Se realizó un análisis factorial múltiple sobre 6 grupos de variables diferentes. Estos grupos se conformaron con 5 variables fisicoquímicas relevadas in situ, 14 variables referidas a minerales y nutrientes, 2 variables bacteriológicas, 11 referidas al hábitat acuático, 6 que expresan la composición del suelo y una indicadora de calidad de agua (Índice BMPS cualitativa). Se buscó determinar relaciones entre los grupos de variables, estudiar la configuración de los sitios al utilizar diferentes grupos de variables simultáneamente y determinar la relación entre los diferentes grupos de variables y la calidad del agua. Se vio una clara distinción de la variable suplementaria indicadora de la calidad del agua con la primer dimensión. Se encontró una relación entre el tipo de suelo de cada sitio y la presencia de iones y nutrientes, entre otras.

Palabras claves: análisis multivariado; análisis a tres vías; calidad de agua

APLICACIÓN DE TÉCNICAS DE METAANÁLISIS PARA LA DETECCIÓN DE EXPRESIÓN DIFERENCIAL DE GENES EN LA CANCROSIS DE LOS CÍTRICOS

LAURA PISKULIC¹, MARCOS PRUNELLO¹ y LUCAS DAURELIO²

¹ ÁREA ESTADÍSTICA Y PROCESAMIENTO DE DATOS, FACULTAD DE CIENCIAS BIOQUÍMICAS Y FARMACÉUTICAS, UNR. ² ÁREA ESTADÍSTICA Y PROCESAMIENTO DE DATOS, FACULTAD DE CIENCIAS BIOQUÍMICAS Y FARMACÉUTICAS, UNR. CÁTEDRA DE FISIOLOGÍA VEGETAL, FACULTAD DE CIENCIAS AGRARIAS, UNL; CONICET.

lpiskulic@yahoo.com

Introducción: Se denomina metaanálisis (MA) al uso de técnicas estadísticas para combinar resultados de experimentos independientes pero referentes al mismo problema. El MA tiene la virtud tanto de aumentar la potencia como de generalizar conclusiones a partir de estudios únicos. En lo que respecta a los experimentos de expresión génica, los cuales miden la cantidad de transcripto o ARN mensajero que se genera a partir de un gen codificado en el ADN, el MA permite aumentar la potencia para detectar genes diferencialmente expresados (DE), es decir aquellos que cambian su expresión en forma significativa al comparar dos o más tratamientos. La interacción entre las plantas y los patógenos, que puede dar como resultado que la planta presente defensa o enfermedad, involucra cambios en la expresión génica de ambos. Las técnicas de análisis de expresión génica como los microarreglos, junto con la disponibilidad de secuencias genómicas para algunas especies vegetales, han permitido un significativo progreso en la caracterización de genes implicados en la interacción planta-patógeno. La producción de cítricos, principal cultivo frutícola en la Argentina, se ve afectada por la cancrosis, enfermedad producida por la bacteria *Xanthomonas citri* pv. *citri* (Xcc). **Objetivo:** Aplicar diferentes técnicas estadísticas de MA sobre experimentos de expresión génica para detectar genes de cítricos que responden a la infección por Xcc. **Metodología:** Se aplicaron 7 métodos de MA basados en la combinación de los valores p obtenidos en 5 experimentos de microarreglos, realizados sobre hoja de *Citrus sinensis* durante la infección por cancrosis. Los métodos aplicados fueron: mínimo valor p, máximo valor p, valor p de orden 3, Fisher, Stouffer, Suma de Rangos y Producto de Rangos. **Resultados:** De los 7167 genes estudiados, 232 resultaron DE por lo menos en 5 de las 7 técnicas aplicadas. De estos, 68 fueron DE en todas las técnicas, y 43 en 6 de ellas. El estudio funcional de este conjunto de genes mediante Singular Enrichment Analysis permitió identificar 23 categorías significativas, entre las cuales se destacan traducción y fotosíntesis, y un número importante de genes relacionados a estrés biótico, regulación y hormonas. **Conclusión:** El uso de técnicas de MA permitió identificar genes DE en 5 experimentos diferentes,

cuyo análisis contribuirá a la interpretación de los mecanismos moleculares involucrados en la interacción planta-patógeno que conduce a la cancrosis de los cítricos. Estos resultados además podrán ser utilizados para buscar alternativas que permitan combatir esta importante enfermedad.

Palabras claves: metaanálisis; expresión génica; microarreglos; interacción planta-p; cancrosis de los cít

ASISTENCIA OBLIGATORIA EN CURSOS DE ESTADÍSTICA PARA ESTUDIANTES DE CIENCIAS BIOLÓGICAS Y QUÍMICAS: ENTRE LA PERCEPCIÓN DE LOS ALUMNOS Y EL LOGRO DE LOS OBJETIVOS CURRICULARES

***LAURA PISKULIC¹, MARCOS PRUNELLO², LILIANA CUCCIARELLI²,
MARÍA BELÉN ALLASIA², SABRINA SILVA QUINTANA², JUAN JOSÉ
IVANCOVICH², LILIANA RACCA² y HEBE BOTTAI²***

¹ÁREA ESTADÍSTICA Y PROCESAMIENTO DE DATOS, FACULTAD DE CIENCIAS BIOQUÍMICAS Y FARMACÉUTICAS, UNIVERSIDAD NACIONAL DE ROSARIO.. ²ÁREA ESTADÍSTICA Y PROCESAMIENTO DE DATOS, FACULTAD DE CIENCIAS BIOQUÍMICAS Y FARMACÉUTICAS, UNIVERSIDAD NACIONAL DE ROSARIO.

lpiskulic@yahoo.com

El conocimiento de los fundamentos del cálculo de probabilidades, del diseño de experimentos y del análisis descriptivo e inferencial son indispensables en múltiples aspectos de la Biología y la Química. La asignatura Estadística en la Facultad de Ciencias Bioquímicas y Farmacéuticas de la Universidad Nacional de Rosario persigue el objetivo de introducir al estudiante en la Estadística, resaltando su papel en la construcción del conocimiento científico, y desarrollar una actitud crítica y reflexiva en el análisis e interpretación de un estudio de investigación. Esta asignatura forma parte del ciclo de formación general de todas las carreras dictadas en la facultad, contando con más de 400 inscriptos. El cursado consiste en clases teórico-prácticas de asistencia obligatoria. La exposición verbal se integra con la presentación de situaciones problemáticas cuya resolución conduce al alumno a buscar información, discutir con sus pares, analizar datos y arribar a conclusiones. El alumno alcanza la condición de “Regular” aprobando un examen práctico o su recuperatorio, “Abandonó” si no se presentó a ninguna actividad evaluativa o “Insuficiente” en otro caso. En 2016 se eliminó la asistencia obligatoria. Se realizó una encuesta a los estudiantes, encontrándose que el 71% estuvo a favor de la no obligatoriedad, el 60% opinó que la misma no modificaría el rendimiento y el 50% que la obligatoriedad no contribuye a llevar la asignatura al día. A pesar de que el 76.8% reconoció que faltar a clase dificulta la regularización, la asistencia mediana disminuyó a 58%, siendo de 81.8% y 82.5% en 2014 y 2015, respectivamente. Al comparar los resultados académicos se observó un peor desempeño de los estudiantes. El porcentaje de regulares fue 46.6% en 2014, 52.1% en 2015 y 34.2% en 2016, mientras que el de insuficientes fue 46.1%, 40.3% y 53.5%. Entre aquellos que asistieron al menos al 80% de las clases, no hubo diferencias en la distribución de la condición final a través de los tres años. Esta experiencia pone de manifiesto una discordancia entre la creencia de los estudiantes sobre los beneficios del cursado y los resultados obtenidos cuando la obligatoriedad es removida, tanto en el porcentaje de asistencia como en el rendimiento académico. Además, resalta la importancia del trabajo presencial y en conjunto entre el docente y los estudiantes para alcanzar los objetivos de la enseñanza-aprendizaje de Estadística en carreras de grado no afines. Como resultado, la asistencia obligatoria a clases fue instaurada nuevamente para el ciclo lectivo 2017.

Palabras claves: enseñanza-aprendizaje; asistencia; regularización

**BÚSQUEDA DE TAMAÑO DE MUESTRA Y DIMENSIÓN
DE UNIDAD MUESTRAL ADECUADOS PARA LA
CONFECIÓN DE MAPAS DE RENDIMIENTO EN
MANZANOS EN EL ALTO VALLE DE RÍO NEGRO Y
NEUQUÉN**

S OCAMPO¹, D FERNANDEZ¹, M CURETTI¹ y S BRAMARDI²

¹INTA. ²UNCOMA.

ocampo.santiago@inta.gob.ar

El mapa de rendimiento es una de las herramientas más utilizadas en agricultura de precisión. Con el objetivo de determinar el tamaño de muestra más apropiado para generar mapas de rendimiento con la menor pérdida de información utilizando la técnica de geoestadística, se relevaron datos de manera exhaustiva de una parcela de árboles de manzano cv. Crisp Pink durante la temporada 2014-2015. Se cosecharon y pesaron todos los frutos de cada planta de una parcela de 9 filas con 61 plantas en un marco de plantación de 4 x 1,5 m. A cada una de las plantas se les asignó una coordenada en el sistema UTM. Para la confección de los mapas se utilizó el software GS+. En un primer estudio se conformaron grupos de plantas adyacentes dentro de cada fila (ej.: 274 grupos de 2 plantas, 183 de 3, etc., hasta 18 grupos de 30 plantas). Se calculó la producción promedio de cada grupo y su coordenada promedio. Para cada tamaño de agrupamiento, se realizó un mapa de rendimiento y se evaluó su capacidad de representación respecto del mapa obtenido con el muestreo exhaustivo de las 549 plantas. La comparación se hizo mediante el coeficiente de correlación r. Se encontró un descenso de la capacidad de representación a medida que el nivel de agrupamiento se incrementa, con un $r=0,98$ si se mapea con los 549 datos, un $r=0,63$ si se mapea con los datos agrupados de a 6 plantas vecinas y un $r=0,30$ si se mapea con los datos agrupados de a 30 plantas vecinas. De este modo, no parece conveniente confeccionar mapas utilizando esta técnica, agrupando más de 6 plantas. En un segundo estudio, de los 549 datos de rendimiento disponibles, se seleccionaron diferentes muestras de manera aleatoria y se confeccionaron mapas geoestadísticos a partir de los datos de cada muestra. Se utilizaron tamaños de muestra de 50, 100, 150 y hasta 450. Para cada caso se comparó la capacidad de representación de los mapas resultantes vs. el mapa obtenido con los 549 datos, mediante el coeficiente de correlación. Como es de esperar, la capacidad de representación se incrementa con el tamaño de muestra, obteniéndose un $r=0,36$ para una muestra de tamaño 50, $r=0,60$ para 150, $r=0,70$ para 250, entre otros. La muestra aleatoria de 150 unidades parecería ser la menor a partir de la cual se logra una representación satisfactoria del rendimiento del cuadro.

Palabras claves: agricultura de preci; geoestadística; pronósticos de rendi; datos geo-referencia

**CENSO NACIONAL DE FAMILIAS CON MIEMBROS CON
DISCAPACIDAD**

LUIS CASTIGLIONI¹, GRACIELA GAZZERA¹ y ROSANA KUCUKBEYAZ²

¹OSDOP. ²UNTREF.

ranush2003@yahoo.com.ar

La obra social OSDOP (Obra Social de Docentes Privados) desde su Gerencia de Prestaciones de Salud decidió realizar un Censo de las familias con miembro/miembros con discapacidad por considerar indispensable conocer la realidad cotidiana de vida y la dinámica de su grupo familiar.

Comenzó este proyecto con una experiencia piloto en la delegación Mar del Plata donde se probó la herramienta censal construida desde la OSDOP. La encuesta relevó información sobre las condiciones de vida de las familias que tiene uno o mas integrante con discapacidad. La cedula censal contiene aspectos cuantitativos y cualitativos. La entrevista fue realizada por una pareja de encuestadores formados en la institución : un Licenciado en Trabajo Social – personal externo- y un Agente Sanitarios –personal interno- La cedula censal presenta dos dimensiones : La familia y el afiliado/s con discapacidad En número total de afiliados a censar fue de 740, llevándose a cabo 660. No contamos con antecedentes sobre la realización de un Censo de estas características lo cual posibilitó concretar desde foja cero, datos que permitan contar con información confiable y útil. Este Censo es el inicio de un Proyecto mayor que permitirá realizar un trabajo más amplio sobre el total de nuestra población de afiliados con discapacidad y sus familias. Fundamentalmente pondrá en evidencia las debilidades y fortalezas de la Obra Social y sus prestaciones. Identificar situaciones invisibilizadas en el hacer diario de la actividad de la OSDOP. Conocer y captar las sugerencias que aporten las familias sobre nuestra calidad de atención. Construir un puente permanente de contacto. Verificar los datos del padrón.

Palabras claves: censo; discapacidad; osdop

DATOS ABIERTOS: PUBLICACIÓN DE DATOS EPIDEMIOLÓGICOS ROSARIO, PERÍODO 2015-2017

VALERIA AGUZZI¹, LAURA BALPARD², MARIELA CAMUSSI², MARIO CHAVERO², ANALÍA CHUMPITAZ², JORGE DIMARCO¹, PATRICIA GIARDINI¹, DIEGO LÓPEZ³, ANDREA MORO², MARÍA EUGENIA PERETTI⁴, CINTIA SILVA² y MARINA SUAREZ²

¹DIRECCIÓN GENERAL DE INFORMÁTICA. SECRETARÍA GENERAL. MUNICIPALIDAD DE ROSARIO.. ²SISTEMA MUNICIPAL DE EPIDEMIOLOGÍA. SECRETARÍA DE SALUD PÚBLICA. MUNICIPALIDAD DE ROSARIO.. ³ÁREA DE SENsoRES REMOTOS. ESCUELA DE AGRIMENSURA. FACULTAD DE CIENCIAS EXACTAS, INGENIERÍA Y AGRIMENSURA. UNIVERSIDAD NACIONAL DE ROSARIO.. ⁴EQUIPO WEB. DIRECCIÓN GENERAL DE COMUNICACIÓN SOCIAL. SECRETARÍA DE GOBIERNO. MUNICIPALIDAD DE ROSARIO..

lbalparda@hotmail.com

Datos abiertos es una iniciativa a nivel mundial cuyo objetivo es la publicación de datos de la gestión pública, en forma completa y oportuna, en el marco de un proceso proactivo, “en y por los canales, medios, formatos y bajo la licencia que mejor facilite su ubicación, acceso, procesamiento, uso, reutilización y redistribución”¹. De este modo, se pretende que la difusión de datos públicos contribuya a fortalecer el “proceso democrático, el desarrollo de políticas públicas basadas en la evidencia, la provisión de servicios públicos, la promoción del desarrollo social, económico, científico y cultural”¹ de una sociedad. En el año 2009, países como EEUU, Gran Bretaña y Canadá se sumaron a esta propuesta. En Argentina, la Constitución Nacional garantiza el principio de publicidad de los actos de gobierno y el derecho a la información pública (artículos 33, 41, 42 y 75 inciso 22); en tanto que, el Decreto 117/2016 del Plan Nacional de Apertura de Datos promulga que las áreas dependientes del Poder Ejecutivo Nacional realicen un detalle de los activos de los conjuntos de datos (datasets) y un cronograma de publicación. La Municipalidad de la ciudad de Rosario reglamentó la creación del programa “Open Data Rosario” (ordenanza 9279/2014). Actualmente, el portal “Rosario Datos” (datos.rosario.gob.ar) publica 510 archivos en formatos abiertos agrupados en 160 datasets de diversas temáticas: territorio, población, movilidad, salud, administración financiera, entre otros. El Sistema Municipal de Epidemiología, adhiriendo a esta política de socialización de datos, publicó en el portal municipal 29 datasets correspondientes a Eventos de Notificación Obligatoria (ENOs), tales como dengue, coqueluche, meningitis, hantavirus, leptospirosis, psitacosis, triquinosis y zika, correspondientes al período 04/01/2015 a 01/07/2017. Los datos, agregados a nivel de ciudad (unidad espacial) y por semana epidemiológica (unidad temporal), se actualizarán semestralmente. Cumplida esta primera etapa, el equipo de trabajo de Epidemiología se plantea superarla, publicando las características de la unidad de análisis original (paciente) con un mayor

nivel de desagregación espacial (adicionando el radio/fracción censal, según el domicilio de residencia del paciente) y temporal (fecha de ocurrencia del caso). En todos los registros se omitirán los datos filiatorios (nombre, apellido, DNI y dirección), respetando el secreto estadístico (Ley 17.622, artículo 10). Esto permitiría a los usuarios disponer de más datos, ampliando las posibilidades de uso de los datos abiertos. 1 Boletín Oficial República Argentina. Buenos Aires [citado 02/08/2017]. Plan de Apertura de Datos, Decreto 117/2016. Disponible en: <https://www.boletinoficial.gob.ar/>.

Palabras claves: datos abiertos; epidemiología; datos agregados-desa

ELABORACIÓN DE INDICADORES PARA LA GESTIÓN PÚBLICA: EL CASO DE LOS INDICES DE OBRAS PUBLICAS DE CÓRDOBA

LAURA ISABEL LUNA, VERÓNICA ARIAS y MARIANA DIAZ

DIRECCIÓN GENERAL DE ESTADÍSTICAS Y CENSO.

lauraisabel.luna@gmail.com

ELABORACIÓN DE INDICADORES PARA LA GESTIÓN PÚBLICA: EL CASO DE LOS INDICES DE OBRAS PUBLICAS DE CÓRDOBA. Luna, Laura Isabel*, Arias, Verónica*, Díaz, Mariana* *Gobierno de la Provincia de Córdoba, Secretaría de Fortalecimiento Institucional, Dirección General de Estadística y Censos lauraisabel.luna@cba.gov.ar, veronica.arias@cba.gov.ar, mariana.diaz@cba.gov.ar **RESUMEN** El presente trabajo tiene como objetivo difundir y poner al alcance de los especialistas y usuarios en general los principales aspectos metodológicos y prácticos vinculados a la construcción de los índices de “Obra Pública” de Córdoba. Es fundamental la calidad, exactitud y oportunidad de esta importante herramienta para la gestión pública. El objetivo principal de estos índices es reflejar la variación de los precios medios de una serie de factores, constituyéndose en exclusivos indicadores de referencia para la redeterminación de Precios de Obras Públicas del Gobierno de la Provincia de Córdoba, tal como lo establece el Decreto 800/16. Cada índice se calcula en base a la variación del precio medio del respectivo insumo, ó bien, como un promedio ponderado de las variaciones de los precios medios de una canasta de insumos. Se relevan precios de 187 productos, que en distinto grado inciden en los indicadores que se publican, y se incluyen dos índices generales, el Índice del Costo de la Construcción de Córdoba y el Índice de Precios al Consumidor de Córdoba, ambos relevados mensualmente. En el presente trabajo se analizan las principales características de este conjunto de índices y se detallan los métodos y procedimientos utilizados para su cálculo, también se presentan los resultados para la serie 2015-2017. Consideramos que la publicación y discusión de los mismos es fundamental para la transparencia, confiabilidad y mejora continua de las estadísticas oficiales. Palabras claves: estadísticas oficiales, índices de obra pública, redeterminación de precios, provincia de Córdoba.

Palabras claves: estadísticas oficial; índices de obra públ; redeterminación de p; provincia de cordoba

ELABORACIÓN DE INDICADORES PARA LA GESTIÓN PÚBLICA: EL CASO DEL FACTOR MANO DE OBRA DE CÓRDOBA

PATRICIA MASERA y MARIANA DIAZ

GOBIERNO DE LA PROVINCIA DE CÓRDOBA, SECRETARÍA DE FORTALECIMIENTO INSTITUCIONAL, DIRECCIÓN GENERAL DE ESTADÍSTICA Y CENSO.

mariana.diaz@cba.gov.ar

El presente trabajo tiene como objetivo difundir y poner al alcance de los especialistas y usuarios en general los principales aspectos metodológicos y prácticos vinculados a la construcción del índice “Factor Mano de Obra” de Córdoba. Es fundamental la calidad, exactitud, y oportunidad de esta importante herramienta para la gestión pública. Este indicador es un insumo para el sistema de redeterminación de precios por reconocimiento de variación de costos para la provisión de bienes o prestaciones de servicios del Régimen de Compras y Contrataciones de la Administración Pública Provincial establecido por Ley N° 10.155 y su Decreto Reglamentario N° 305/2014. A través del Decreto 1160/16 se designó a la Dirección General de Estadística y Censos como el organismo encargado de relevar, calcular y publicar mensualmente la evolución de los cambios en los precios medios de los insumos más relevantes utilizados y consumidos en los distintos tipos de contrataciones de Bienes y Servicios llevadas a cabo en la provincia de Córdoba. Uno de estos indicadores es el “Factor Mano de Obra” el cual se desagrega en “sector privado registrado” y “sector público provincial”. Su objetivo principal es reflejar la variación de los salarios medios en las principales actividades productivas de la Provincia. En lo que refiere al sector privado registrado, el cálculo contempla aquellas actividades económicas más relevantes en la provincia, según su masa salarial bruta. Asimismo se incluyeron, independientemente de la masa salarial, aquellas actividades más representativas en las contrataciones relacionadas a la compra y adquisición de bienes y servicios por parte del Estado Provincial. En el presente trabajo se analizan las principales características del índice Factor Mano de Obra Córdoba y se detallan los métodos y procedimientos utilizados para su cálculo, también se presentan los resultados para la serie 2016-2017. Consideramos que la publicación y discusión de los mismos es fundamental para la transparencia, confiabilidad y mejora continua de las estadísticas oficiales

Palabras claves: estadísticas oficial; índice de salarios; provincia de córdoba; redeterminación de precios

EMPLEO DE MODELOS DE MEZCLA CON DISTRIBUCIONES BETA PARA LA IDENTIFICACIÓN DE GENES DIFERENCIALMENTE METILADOS EN UN ESTUDIO PAN-CÁNCER

MARCOS PRUNELLO¹ y OLIVERT GEVAERT²

¹FACULTAD DE CIENCIAS BIOQUÍMICAS Y FARMACÉUTICAS, UNIVERSIDAD NACIONAL DE ROSARIO; CONICET. ²BIOMEDICAL INFORMATICS RESEARCH, DEPARTMENT OF MEDICINE, STANFORD UNIVERSITY.

mprunello@fbiqyf.unr.edu.ar

La metilación del ADN es la transferencia de un grupo metílico en sitios CpG en la cadena del ADN que produce importantes efectos regulatorios. Se conoce que tanto la hipo- como la hiper-metilación juegan un rol significativo en cáncer por lo cual resulta de interés identificar genes con metilación anormal. A través de un algoritmo que modela la metilación a través del ajuste de modelos de mezcla para distribuciones beta y que ajusta la relación entre la metilación y la expresión génica con modelos lineales, identificamos genes diferencialmente metilados y transcripcionalmente predictivos en un estudio pan-cáncer consistente de 26 tipos de tumores y más de 9000 muestras. Encontramos 146 genes con metilación diferencial en muestras de cáncer con respecto a muestras normales. Characterizamos a cada muestra de acuerdo con su perfil de metilación para estos genes y a través de la técnica de consensus clustering identificamos 15 conglomerados de pacientes. Algunos grupos están principalmente constituidos por muestras del mismo cáncer o de cánceres relacionados, mientras que otros están formados por muestras de distintos tipos de tejidos reflejando similitudes en sus perfiles de metilación. Los conglomerados asociados a tumor cerebral presentaron el mayor número de genes hiper-metilados y los relacionados a cáncer de ovarios y tiroides, el menor. La supervivencia resultó significativamente diferente entre los conglomerados pero también entre muestras

del mismo tejido que fueron agrupadas en distintos conglomerados. Analizamos también variables clínicas de los pacientes y encontramos diferencias significativas en cuanto a la edad, hábito de fumar, género, estado de la enfermedad, grado del tumor e histología. Para algunos cánceres, la separación de muestras de un mismo tejido en distintos conglomerados está asociada a tipificaciones basadas en expresión génica. Empleamos SAM (Significance Analysis of Microarrays) para identificar genes sobre-expresados en cada conglomerado, los cuales fueron posteriormente analizados con GSEA (Gene Set Enrichment Analysis). Varios conjuntos de genes conocidos por estar regulados por metilación mostraron enriquecimiento y en muchos casos el conjunto de genes sobre-expresados en un conglomerado particular resultó estar vinculado a conjuntos de genes previamente asociados al tipo de cáncer dominante en ese conglomerado. Este análisis revela nuevas similitudes entre tejidos malignamente transformados basadas en patrones de metilación comunes y puede ser útil para re describir aspectos del cáncer en función de la metilación en lugar de sitio de origen.

Palabras claves: modelos de mezcla; distribución beta; conglomerados; metilacion de adn; cáncer; expresión génica

ESTIMACIÓN DEL TRASVASE DE VOTOS ENTRE PARTIDOS APLICACIÓN DE UN MODELO MATEMÁTICO A LOS RESULTADOS DE LA ELECCIONES DE LA PROVINCIA DE SANTA FE AÑOS 2011-2015

GREGORIO GARCÍA y JOSÉ ALBERTO PAGURA

ESCUELA DE ESTADÍSTICA. FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA. UNIVERSIDAD NACIONAL DE ROSARIO.

jpagura@fcecon.unr.edu.ar

Un tema relevante a la hora del análisis de los resultados de las elecciones, es el desplazamiento de votos de un partido hacia otro conocido como “trasvase de votos”. Quienes se interesan en conocer el comportamiento de este fenómeno, por lo general buscan la información por medio de encuestas. Además del costo que representan estos estudios y el tiempo que transcurre hasta la obtención de los resultados finales, puede haber falta de respuesta, o respuestas falsas por olvido u omisión. En el año 2014, el académico valenciano Rafael Romero Villafranca publicó un procedimiento matemático para estimar el trasvase de votos entre partidos a partir de los resultados electorales en diferentes unidades territoriales de una región de interés, aplicándolo con éxito en diferentes elecciones europeas hasta autonómicas. Dicho método se basa en programación lineal y se puede aplicar sobre los resultados electorales de una elección y la anterior. En este procedimiento es crucial la hipótesis de homogeneidad, esto es, la probabilidad de pasar de votar un partido a votar otro, es la misma para todas las unidades territoriales. Otra hipótesis necesaria para ejecutar el proceso de optimización, es el establecimiento de un “umbral de lealtad” es decir, la proporción de votantes que en ambas elecciones votan al mismo partido. Un aparente problema es la forma en la que se consideran aquellos votantes que se incorporan a la elección actual y aquellos que votaron en la elección anterior y ya no forman parte del padrón electoral, cuya solución es la creación de dos partidos ficticios llamados “altas” y “bajas”. Para la inclusión de votos en blanco y nulos se emplea el mismo recurso. En el presente trabajo, se muestran las estimaciones de trasvase de votos entre los años 2011 y 2015 en las elecciones a gobernador, para cada departamento de la provincia de Santa Fe tomando como unidad territorial cada localidad. El procedimiento fue implementado computacionalmente en R.

Palabras claves: movilidad electoral; programación lineal; simplex

ESTIMACIÓN Y PRONÓSTICOS DE TASAS BÁSICAS EN EL AGLOMERADO GRAN ROSARIO UTILIZANDO MÉTODOS DE SUAVIZADO EXPONENCIAL PROBABILÍSTICOS

*SABRINA SILVA QUINTANA¹, LUCÍA ANDREOZZI² y MARÍA TERESA
BLACONÁ²*

¹ÁREA ESTADÍSTICA Y PROCESAMIENTO DE DATOS. FACULTAD DE CIENCIAS BIOQUÍMICAS Y FARMACÉUTICAS. UNIVERSIDAD NACIONAL DE ROSARIO. ROSARIO, ARGENTINA. ²FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA. UNIVERSIDAD NACIONAL DE ROSARIO. ROSARIO, ARGENTINA.

sabrinasilvac@gmail.com

En estudios de mercado laboral, uno de los principales intereses es identificar aquellos grupos que tienen mayores inconvenientes a la hora de desempeñarse, lo cual puede lograrse a través del análisis de las condiciones de ocupación de las personas -si consiguen o buscan trabajo o si están desocupados- y, consecuentemente, llevar a cabo políticas activas de empleo para enfrentarlos y mejorar la situación actual. En este trabajo, se estudian las series de los principales indicadores del mercado de trabajo para Rosario y la región: tasas de Actividad, Empleo y Desocupación; calculadas de manera general, por género y por tres grandes grupos de edad definidos como: Jóvenes (18 a 24 años), Adultos (25 a 45 años) y Adultos Mayores (46 a 65 años). El objetivo de este trabajo es analizar el comportamiento de las tasas trimestrales en el pasado para pronosticar valores futuros. Se proponen a tal fin, los Modelos de Espacio de Estado de Innovación de Hyndman (Hyndman, Koehler, Ord y Snyder, 2008) y los modelos ARIMA. En cada serie, se ha optado por el modelo que mejor ajusta los datos presentados, en otras palabras el modelo que exhibe menor AIC (AKAIKE) entre todos los posibles y se busca detectar aquellos grupos que necesitan mayor atención para acceder a un empleo. Una medida del error de pronóstico de cada modelo es el MAPE "Mean Absolute Percentage Error" de capacidad predictiva, cuando el valor del mismo sea del 10 % o menos, se concluirá que dicho modelo es óptimo para pronosticar. En el 27,78% de las tasas los modelos de Hyndman no detectaron estacionalidad que los modelos ARIMA sí. Cuando la capacidad predictiva de un modelo fue aceptable, la del otro modelo también lo fue y viceversa. Para las tasas de actividad y empleo se obtienen buenos ajustes, mientras que para las tasas de desocupación los modelos no representan adecuadamente los datos y los pronósticos resultan no satisfactorios, esto mismo puede deberse a la gran variabilidad que posee dicha tasa. Por otra parte los grupos donde es necesario centrarse para mejorar su situación laboral son los jóvenes de 18 a 24 años y el género femenino.

Palabras claves: series; tiempo; tasa; actividad; empleo; desocupación

ESTUDIO COMPARATIVO SOBRE HÁBITOS ALIMENTARIOS Y ACTIVIDAD FÍSICA DE ALUMNOS DE LA LICENCIATURA EN NUTRICIÓN, BIOQUÍMICA Y BIOTECNOLOGÍA DE LA UNIVERSIDAD NACIONAL DEL LITORAL

EUGENIA EMILIA BERTA y OLGA BEATRIZ AVILA

DEPARTAMENTO DE MATEMATICA-FACULTAD BIOQUÍMICA Y CIENCIAS BIOLÓGICAS-UNL

berta.eugenia@hotmail.com

Introducción: La etapa universitaria es un periodo decisivo en la consolidación de un estilo de vida saludable, y es de esperar que cuando mayor es la información en nutrición, mejor son los hábitos

alimentarios. Sin embargo, a medida que el estudiante adquiere autonomía para decidir comidas y horarios, los factores sociales, culturales y económicos, además de las preferencias alimentarias, van a contribuir al establecimiento del patrón de consumo alimentario, no traduciéndose estos conocimientos en el seguimiento de una dieta equilibrada. El objetivo de este trabajo fue describir los hábitos alimentarios y de actividad física (AF) de estudiantes y su relación con la carrera elegida. Material y método: Estudio transversal realizado el presente año, a través de un cuestionario anónimo autoadministrado a 164 estudiantes del segundo año las carreras de Lic. en Nutrición, Bioquímica y Biotecnología, pertenecientes a la Facultad de Bioquímica y Ciencias Biológicas, con preguntas cerradas acerca de los hábitos y patrones de consumo alimentario, la actividad física y sus factores de influencia. Resultados: La muestra estuvo representada por un 77% de mujeres y 23% hombres, con edad promedio de 20,7 años ($DE = 2,46$). Se observó que la mayoría presenta conductas saludables en las áreas evaluadas, encontrándose una diferencia estadísticamente significativa entre carreras en relación a la realización diaria del desayuno ($p=0,043$), de las cuatro comidas al día ($p=0,009$) y de la incorporación de azúcar al mate en quienes lo consumen ($p=0,014$), resultando los estudiantes de nutrición quienes presentaban los patrones más saludables. Además, en relación a los motivos de elección alimentaria, el valor nutritivo de los alimentos fue un factor a tener en cuenta preferentemente por los estudiantes de Nutrición tanto para el desayuno ($p=0,003$), almuerzo ($p=0,007$) y merienda ($p=0,000$), no así en la cena momento en el cual un mayor porcentaje de alumnos la realizan en familia, dedicándole más tiempo respecto a las demás comidas. En relación a la AF, no se observaron diferencias entre carreras respecto a los minutos semanales realizados, pero sí entre el tipo de actividad elegida, siendo la caminata la preferida entre estudiantes de nutrición ($p=0,003$). Conclusión: es recomendable propiciar la educación en alimentación saludable y actividad física que beneficien a toda la comunidad universitaria, a través de proyectos de salud y, preferentemente, dentro de los planes curriculares de todas las carreras como parte de una formación integral.

Palabras claves: alimentación; actividad física; estudiantes universi

EVALUACIÓN DE PAPA POS COSECHA MEDIANTE MODELOS MIXTOS LINEALES

NELIDA DEL VALLE ORTIZ y ELIZABETH L. VILLAGRA

UNIVERSIDAD NACIONAL DE TUCUMÁN.

ortiznelidadelvalle@gmail.com

El correcto equilibrio nutricional recibido durante el cultivo de papa, incide en el potencial de conservación del tubérculo en almacenaje. El balance de nutrientes afecta brotación, rendimientos, respiración e incide en la pérdida de agua por lo cual, afecta directamente la calidad de los tubérculos. La biofertilización constituye un método nutricional alternativo y ofrece perspectivas para un manejo sustentable. El objetivo de este trabajo es contrastar efectos de la fertilización química estándar con el manejo combinado de fertilización química y biofertilización sobre poscosecha de tubérculos de papa-consumo producidos en el Pedemonte tucumano. En este estudio se aplicaron siete tratamientos incluido un control. Los tubérculos se trataron en laboratorio bajo condiciones homogéneas, sin embargo, alguno se perdieron en el tiempo por putrefacción de manera aleatoria. La variable respuesta fue el peso de los tubérculos registrados durante diez observaciones en el tiempo con intervalos de diez días y las variables aleatorias de predicción fueron los días de observación posteriores a la cosecha y la referida a los tratamientos aplicados al cultivo. Se estima que podría tratarse de error de medida en el ensayo por la pérdida de tubérculos. Los datos se evaluaron mediante el ajuste de modelos mixtos lineales para datos continuos mediante el paquete estadístico nlme de R. Luego se sometió a la evaluación de modelación de variables con error de medida mediante la inferencia del método de simulación extrapolación SIMEX para variable continua. Se obtuvo diferencias entre

los tratamientos utilizados. Se registraron algunas leves diferencias entre errores estándares de los estimadores. Se concluye que el ajuste de modelos mixtos lineales fue apropiado.

Palabras claves: modelo mixto lineal; simex; r

FACTORES DETERMINANTES EN LA PERCEPCIÓN DE LOS EGRESADOS EN CIENCIAS ECONÓMICAS EN RELACIÓN A SU CONSEJO PROFESIONAL APLICANDO UN MODELO DE ECUACIONES ESTRUCTURALES

MARIANA VERÓNICA GONZALEZ¹, ROCÍO M. CERINO², FERNANDO A. PRONELLO² y FRANCO D. VICO²

¹INSTITUTO DE ESTADÍSTICA Y DEMOGRAFÍA. FACULTAD DE CIENCIAS ECONÓMICAS. UNIVERSIDAD NACIONAL DE CÓRDOBA. ARGENTINA.. ²FACULTAD DE CIENCIAS ECONÓMICAS. UNIVERSIDAD NACIONAL DE CÓRDOBA. ARGENTINA..

gonzalezmariana77@gmail.com

Esta investigación plantea una aproximación exploratoria a la vez que innovadora, construyendo un modelo integrado de relaciones para explicar la percepción de los egresados en Ciencias Económicas de la Provincia de Córdoba con relación a su Consejo Profesional (CPCE). Los Consejos Profesionales, como entidades de derecho público que nuclean a los egresados, reglamentan y ordenan el ejercicio de las profesiones, necesitan identificar los factores que determinan la decisión de los profesionales de incorporarse o permanecer en la organización. En efecto, la consolidación de relaciones estables entre cualquier tipo de organización y sus principales clientes se ha convertido en una herramienta imprescindible para garantizar la supervivencia de dichas organizaciones. En este sentido, interesa analizar la Imagen de CPCE de Córdoba, entendida como la manera en que el profesional percibe a la organización, de acuerdo a sus experiencias y la Valoración de la Matrícula, exigida para el ejercicio de la profesión. La muestra estuvo formada por 533 egresados de universidades públicas y privadas de la Provincia de Córdoba, en carreras vinculadas a las Ciencias Económicas. Del total de profesionales encuestados, el 50,2% se encontraba matriculado en el CPCE y el resto no, aunque de éste segundo grupo un 23% estaba considerando matricularse. Se aplicaron modelos de ecuaciones estructurales (structural equations models SEM), con dos variables latentes o constructos: Imagen del CPCE y Valoración de la Matrícula. Luego de varias corridas de los modelos se estimó, por Máxima Verosimilitud, un ajuste con siete indicadores significativos para cada constructo. En el caso de la variable latente Valoración de la Matrícula aparece como factor determinante que tiene un fin recaudatorio. En la Imagen, por su parte, influyen significativamente la mejora en las condiciones de la profesión y el hecho de que la pone en valor frente a la sociedad, entre otras. Se identificó, además, una correlación estadísticamente significativa entre las dos variables latentes, como así también correlaciones entre algunas indicadoras, que se incluyeron en el modelo a fin de mejorar el ajuste. Los indicadores de bondad de ajuste RMSEA, SRMR y CFI resultaron adecuados. Finalmente, se realizaron ajustes por sexo, concluyendo que no hay diferencias significativas, estadísticamente, entre hombres y mujeres en el sentido de la Imagen y Valoración para los indicadores considerados.

Palabras claves: ecuaciones estructurales; imagen; valoración

GAMMA MIXED MODEL FOR LONGITUDINAL DATA ANALYSIS: MEMORY EVALUATION OF RATS WITH CEREBRAL ISCHEMIA COMBINED WITH DIABETES

M.H.D. RIBEIRO¹, M. V. O. PERES², A. NUNES³, R.M.W. DE OLIVEIRA³, H. MILANI³ y I. PREVIDELLI⁴

¹DEPARTMENT OF MATHEMATICS, FEDERAL TECHNOLOGICAL UNIVERSITY OF PARANÁ, PATO BRANCO, PARANÁ, BRAZIL.

²DEPARTMENT OF SOCIAL MEDICINE, UNIVERSITY OF SAO PAULO, RIBEIRAO PRETO MEDICAL SCHOOL, BRAZIL. ³DEPARTMENT OF PHARMACOLOGY AND THERAPEUTICS, STATE UNIVERSITY OF MARINGÁ, MARINGÁ, PARANÁ, BRAZIL. ⁴DEPARTMENT OF STATISTICS, STATE UNIVERSITY OF MARINGÁ, MARINGÁ, PARANÁ, BRAZIL.

mribeiro@utfpr.edu.br

Studies with repeated measures have gained prominence in the scientific scenario in recent years, mainly in the areas of health and biological sciences. This is due to the interest of researchers evaluate repeatedly one or more experimental units in order to know the effectiveness of a treatment or the evolution of a patient through an intervention. In this sense, a methodology that can be used to model data thus characterized, when the distribution of the response variable belongs to the exponential family, is that of the generalized linear mixed models, whose focus is on accommodating the inter and intra individual variations of the longitudinal measurements. The estimation of the vector of fixed parameters of this model as well as the components of the variance-covariance matrix will be done through the maximization of the marginal likelihood function obtained through integration under the random effects . However, this integration have not analitical solution. Thus is necessary to use numerical integration methods, which provide approximate solutions to given integrals . Thus, in this study a mixed gamma model with logarithmic link function and random effects with normal distribution was used to evaluate longitudinal behavioral data, by means latency response variable. These measure is a positive real number, and data are assymetric. So the gamma distribution is a candidate to modeling data with these characteristics. In this case, the effects of cerebral ischemia, was induced by chronic cerebral hypoperfusion combined with diabetes, on the performance of long-term retrograde memory were studied in rats. The R software was used to fit the model. Based on the results obtained, the proposed model presented good fit (evaluation by means half-normal plot) and accommodated the correlation inherent to the data. Residual analysis was satisfactory (fitted values vs earson Residual). Through influence analysis influential points was not observed. Using this modelating was possible identify that physiologicly normal animals had better results than other animals. Therefore, this methodology becomes an alternative to the use of frequently used traditional methods, which disregard the effect of longitudinal dependence and may make inferences developed inappropriate.

Palabras claves: correlation; gamma distribution; random effect

IDENTIFICACIÓN DE ZONAS PARA MANEJO DIFERENCIADO EN FRUTICULTURA DE PRECISIÓN EN FRUTOS DE PERA EN EL ALTO VALLE DE RÍO NEGRO Y NEUQUÉN

S OCAMPO¹, D FERNANDEZ¹, M CURRETTI¹ y S BRAMARDI²

¹INTA. ²UNCOMA.

ocampo.santiago@inta.gob.ar

El mapa de rendimiento es una de las herramientas más utilizadas en agricultura de precisión. Con el objetivo de diferenciar diferentes zonas de manejo se trabajó en una parcela de 1.8 Ha de pera

cv. Williams. La cosecha se realizó durante cinco temporadas (2014 a 2017) con una plataforma autopropulsada que cuenta con un sistema de pesado de la fruta. La parcela cuenta con 57 filas con 36 plantas y está plantada a 4x2 m. Los datos se colectaron acumulando el peso de la fruta de 12 plantas vecinas (3 datos por fila). Cada grupo de 12 plantas conformó una unidad estadística (UE) a la que se le asoció un dato de producción acumulada y una coordenada en el sistema UTM (168 UE). A su vez se conformaron grupos de UE vecinas para establecer diferentes sectores dentro del cuadro (por ejemplo, agrupando las 168 UE de a 4, se establecen 42 sectores dentro del cuadro). Se estudiaron las diferencias en la producción a lo largo de las cuatro temporadas y en los diferentes sectores del cuadro, y se evaluó la presencia de interacción entre la temporada y el sector. El estudio se llevó a cabo utilizando análisis de la varianza sobre la producción, utilizando el año y el sector de la parcela como factores fijos del modelo. También se utilizó un término de interacción año-sector en el modelo. Si para cierto sector se evalúa la relación entre su nivel de producción respecto al promedio de toda la parcela y se compara esta relación a lo largo de diferentes años, se puede evaluar la estabilidad o inestabilidad de cada sector de la parcela a lo largo de los años. Es posible medir si existe inestabilidad de los sectores a través del término de interacción del análisis de la varianza. No se encontraron interacciones significativas para los años 2015 a 2017, es decir que los sectores mostraron comportamiento estable ($P > 0,5$). Se vio un marcado efecto del año de cosecha y del sector de la parcela sobre la producción (ambos $P < 0,001$). Cuando se incorporó el año 2014, solo algunos de los sectores mostraron inestabilidad ocasionando que la interacción sea ligeramente significativa ($P < 0,003$). Con todo esto fue posible identificar sectores de baja, media y alta productividad y establecer regiones dentro de la parcela para realizar prácticas culturales diferenciadas en el futuro.

Palabras claves: geoestadística; pronósticos de rendimiento; datos geo-referencia; análisis de la varianza

IMPLEMENTACION DE HERRAMIENTAS PARA LA PREDICCION GENÓMICA EN INFOSTAT

JULIO A DI RIENZO¹, PABLO A PACCIORETTI¹ y FERNANDO CASANOVES²

¹FACULTAD DE CIENCIAS AGROPECUARIAS, UNIVERSIDAD NACIONAL DE CÓRDOBA, ARGENTINA. ²UNIDAD DE BIOMETRÍA/CENTRO AGRONÓMICO TROPICAL PARA LA ENSEÑANZA Y LA INVESTIGACIÓN (CATIE), COSTA RICA.

dirienzo@agro.unc.edu.ar

Se presenta un módulo para la predicción genómica implementado en el software estadístico InfoStat (www.infostat.com.ar). El módulo está orientado a la predicción de fenotipos a partir de marcadores SNPs. El módulo incluye un procedimiento para la lectura de SNPs desde archivos en formato .vcf, la selección de marcadores teniendo en cuenta la proporción de datos faltantes y la frecuencia del alelo menos frecuente e implementa un algoritmo para la imputación de datos faltantes. Los SNPs se evalúan en un conjunto de genotipos cuyos fenotipos han sido evaluados experimentalmente. InfoStat cuenta con herramientas de modelación avanzada que permiten obtener estimaciones del fenotipo esperado mediante el ajuste de modelos mixtos o generalizados mixtos en los diversos diseños utilizados en la fenotipificación. La predicción de valor genotípico a partir de los SNPs se implementa a través de tres técnicas de predicción: Ridge Regression, Random Forest y PLS. El procedimiento de calibración guarda una copia del predictor para su posterior utilización. Una vez elegido el algoritmo de predicción, el módulo incluye un procedimiento para predecir fenotipos a partir de un genotipo dado. El procedimiento valida los SNPs disponibles en el predictor y en la muestra a predecir. El módulo permite evaluar el desempeño de las técnicas de predicción por validación cruzada. Para ello el algoritmo divide aleatoriamente (y reiteradamente) el conjunto de genotipos disponibles en muestras de entrenamiento (70%) y validación (30%), para tener una estimación de la capacidad predictiva y su precisión. Los algoritmos numéricos dependen de R y de las librerías vcfr, pls, randomForest y ridge.

Palabras claves: snp; ridge regression; random forest; pls

INTERPRETACIÓN Y LECTURA DE GRÁFICOS ESTADÍSTICOS EN LOS INGRESANTES A LOS PROFESORADOS DE BIOLOGÍA Y MATEMÁTICA

CELINA CORRÍAS, ADRIANA D'AMELIO y LAURA ROSSI

UNIVERSIDAD NACIONAL DE CUYO.

estat0@hotmail.com

Interpretación y lectura de gráficos estadísticos en los ingresantes a los profesorados de Biología y Matemática. Prof.Celina Corrías, Mg.Adriana D'Amelio, Lic. Laura Rossi Universidad Nacional de Cuyo celina.corrias@gmail.com,adriana.damelio@fce.uncu.edu.ar, laura.rossi@fce.uncu.edu.ar Los gráficos estadísticos son muy útiles para resumir y comunicar información en forma precisa, permiten visualizar con mayor claridad y facilidad conceptos o relaciones entre variables muy complejas. Existe actualmente consenso entre los estadísticos e investigadores en educación acerca de la importancia de que un ciudadano pueda leer de forma crítica y comprensiva los gráficos estadísticos que proporcionan los medios de comunicación, internet o ámbitos de trabajo.. Arteaga; Batanero; Díaz; (2009) y Monroy Santana (2007), coinciden en que se suele caer en la falsa apreciación de que la comprensión de gráficos estadístico es una tarea sencilla que no requiere de una especialización. Según estos autores la lectura e interpretación de gráficos estadísticos es una habilidad que no se adquiere de forma espontánea debido a su gran complejidad. Concluyeron que en muchos casos los futuros profesores carecían de los conocimientos matemáticos suficientes para la lectura de dichos gráficos. El objetivo principal de este trabajo, es identificar el nivel de comprensión de gráficos estadísticos de ingresantes a los profesorados de educación media en Biología y Matemática del Instituto de Educación Superior de Formación Docente y Técnica N° 9-002 Tomás Godoy Cruz. Se plantea como hipótesis que los alumnos presentarán dificultades en el reconocimiento de los elementos que componen un gráfico estadístico, en la comparación e interpretación de valores y en la realización de proyecciones e inferencias. Los elementos de análisis se tomarán del instrumento aplicado a la muestra que consiste en la resolución de un cuestionario. El análisis y discusión de resultados tendrá como marco teórico los niveles de comprensión de gráficos estadísticos de Curcio (1989). La categorización en niveles de comprensión sobre gráficos estadísticos de los alumnos proporcionará información sobre el estado actual del conocimiento de dichos alumnos, dando lugar a la reflexión sobre múltiples aspectos: estrategias de enseñanza, tipo de actividades proporcionadas a los alumnos, instrumentos didácticos, tipo de evaluación, etc. En consecuencia el presente proyecto podrá sentar bases para futuras investigaciones y proyectos educativos que favorezcan la alfabetización estadística de los alumnos. La sociedad moderna produce y maneja gran cantidad de datos útiles para la toma de decisiones de los gobiernos, empresas y ciudadanos en general. Por lo tanto desarrollar capacidades para leer y comprender gráficos estadísticos es una necesidad social. Palabras claves. Lectura de gráficos. Alfabetización.Educación. Áreas Temáticas Educación Estadística

Palabras claves: lectura de gráficos; alfabetización; educación

LA ORQUESTA ESTADISTICA AL SERVICIO DE LA PSICOMETRIA

MARIA PIA CAUMONT¹ y SILVIA MARCELA MARIANO²

¹SAE. ²SAE - IASI.

mariapiacaumont@yahoo.com.ar

Se presenta la descripción del análisis estadístico de dos técnicas psicométricas que conjuntamente estudian el significado de las palabras. Objetivo: Describir los recursos estadísticos utilizados para dos instrumentos psicométricos. Materiales y métodos: Se recolectó información en dos instituciones públicas midiendo el significado de dos conceptos para obtener los sentidos tanto denotativos como connotativos de los mismos a través de dos instrumentos psicométricos: Redes Semánticas Naturales y Diferencial Semántico. La primera pide al participante que definan con la mayor precisión posible una palabra estímulo, a través de otra palabras con que la relaciona y se puede observar la riqueza de los conceptos, su densidad y las relaciones. Con la segunda se le pide a los sujetos que ubique un concepto en una serie de escalas cuyos extremos están marcados por adjetivos bipolares. Se mide esa significación a partir de la situación del concepto del objeto analizado en un espacio semántico de dimensiones valorativas. Resultados: A través de métodos estadísticos se desarrolla la presentación de los resultados obtenidos. Por medio de la Red Semántica Natural se mide la riqueza de la red, peso semántico de la definidora, distancia semántica, densidad de la red, se usa recuento, porcentuales, ponderaciones, medianas, test de diferencia de proporciones y Chi cuadrados. Por medio del Diferencial Semántico la "significación" del término para una determinada persona será dada por el perfil resultante en las diferentes escalas de adjetivos, que se reducen a menor cantidad de dimensiones a través del puntaje factorial. Se crea 3 factores: Evaluativo, Potencial y Actividad. Se mide Polarización que es la distancia entre el punto neutral del espacio del diferencial semántico y el concepto particular considerado. Se obtienen medias y medianas de los resultados. Para medir las reacciones asociadas con el estímulo se emplea un análisis factorial. En ambos casos apoyados por representaciones gráficas que sintetizan y facilitan la lectura de la información. Conclusiones: Con el apoyo de técnicas Estadísticas, se hace un aporte al campo de la Psicometría. Relacionando los dos instrumentos permiten el estudio del significado de los conceptos, estableciendo grados de semejanza o disparidad entre ellos y a partir de esta información se puede localizar grupos de entrevistados con perfiles análogos y relacionarlos en base a determinadas características sociales o personales con las respuestas a otras cuestiones.

Palabras claves: psicometria; diferencial semantic; redes semanticas nat

MODELACIÓN DE LA EMERGENCIA DE PLÁNTULAS EN SITUACIÓN DE DESERTIFICACIÓN EN EL OESTE DE RÍO NEGRO: UN PROBLEMA DE SOBREDISPERSIÓN

GUSTAVO GIMÉNEZ¹, DANIEL ZÚÑIGA², M FERNANDA LÓPEZ ARMENGOL³ y NATALIA RUBIO¹

¹DEPARTAMENTO DE ESTADÍSTICA, FACULTAD DE ECONOMÍA, UNIVERSIDAD NACIONAL DEL COMAHUE. ²FACULTAD DE CIENCIAS DEL AMBIENTE Y LA SALUD, UNIVERSIDAD NACIONAL DEL COMAHUE. ³CENTRO DE INVESTIGACIONES EN TOXICOLOGÍA AMBIENTAL Y AGROBIOTECNOLOGÍA DEL COMAHUE (CITAAC), INSTITUTO DE BIOTECNOLOGÍA AGROPECUARIA DEL COMAHUE (IBAC), FACULTAD DE CIENCIAS AGRARIAS, UNIVERSIDAD NACI.

gustavo.gimenez@faea.uncoma.edu.ar

En ecología es frecuente encontrar datos generados a partir de conteos, emergencia de plántulas, frecuencias, abundancia de especies o carga de parásitos. El método frecuentemente empleado para modelar estos datos es asumir que los mismos se aproximan a una distribución Poisson y especificar modelos estadísticos acordes a esta distribución. Sin embargo, un problema persistente con el modelo Poisson es que los datos exhiben una marcada sobredispersión, donde la varianza de la variable de respuesta es mayor a la esperanza, resultando un ajuste pobre de los datos. Esta situación se observó en los datos del presente trabajo: emergencia de plántulas registrada en un campo desertificado en el Oeste de Río Negro durante un año (septiembre de 2013 al 2014) con y sin pastoreo ovino. La emergencia se evaluó sobre dos transectas de 25 m de longitud en un área excluida al pastoreo y dos transectas de igual longitud en sitios pastoreados. Las plántulas se muestrearon mediante un cuadro de 0,1 m² que se situó sobre el terreno a intervalos regulares de 2

m en cada transecta, las plántulas emergidas se contabilizaron y clasificaron de acuerdo al criterio de Raunkiaer que se basa en agrupar a las especies vegetales según su biotipo. Asimismo, se obtuvieron los valores de humedad del suelo en superficie como en profundidad. Inicialmente, se ajustó un modelo Poisson mostrando una marcada sobredispersión donde la devianza superaba ampliamente los grados de libertad del modelo, el test de sobredispersión indicaba un amplio rechazo. Luego, se utilizó un modelo quasi-poisson y un modelo binomial negativo. El modelo que permitió ajustar adecuadamente los datos y contemplar la sobredispersión fue un modelo binomial negativo mixto considerando el efecto transecta y recuadro como efectos aleatorios. Se arribó a este modelo con una devianza de 53. 43 y donde el test de ajuste no fue rechazado (p-valor: 0.46), se comprobó a partir del test de sobredispersión que bajo este modelo no resultaba significativa (p-valor= 0,49). Los residuos del modelo se mostraron satisfactorios, sólo el gráfico residuos versus predichos indicaba una leve falta de ajuste en valores de emergencia bajos. Luego, mediante pruebas de cociente de verosimilitud y valores de Akaike, se dedujo que no existió efecto del pastoreo, aunque se destaca el efecto de los biotipos así como también un importante efecto del mes y del año, la humedad del suelo pudo verse confundido con el efecto mes y año no resultando significativa.

Palabras claves: conteo; modelos generalizado; binomial negativa

MODELOS ESTADÍSTICOS EN ENSAYOS MULTIAMBIENTALES PARA EVALUAR ASOCIACIÓN GENÓMICA

*MARÍA ANGÉLICA RUEDA CALDERÓN, CECILIA BRUNO y MÓNICA
BALZARINI*

angelica8511@gmail.com

En los ensayos agrícolas multiambientales interesa la estimación de la contribución de los efectos de genotipo (G), ambiente (A) e interacción genotipo-ambiente ($G \times A$) en la variación de caracteres cuantitativos. El modelo lineal mixto (MLM) de componentes de varianza es comúnmente usado para este fin. Con la creciente disponibilidad de datos de marcadores moleculares, el efecto global de G puede ser expresado como un modelo aditivo de los efectos de múltiples marcadores y, en consecuencia, también es necesario modelar la interacción $G \times A$ desde la información molecular. Debe considerarse también que en estudios de asociación genómica (Genome-wide association study, GWAS) es común que los individuos de la población bajo análisis se encuentren emparentados genéticamente. Esta relación puede estimarse con coeficientes de coancestría a través del conocimiento del pedigree o bien desde los mismos datos moleculares. El objetivo de este trabajo fue comparar estrategias estadísticas para la estimación de modelos asociación en presencia de interacción $G \times A$. Las estrategias de análisis son comparadas con el MLM de efectos global de G e interacción, cuyas componentes G, $G \times A$ son aleatorias y A es fija. Se trabajó con un conjunto de datos público conformado por 599 genotipos de trigo, genotipados con 1279 SNP y evaluados fenotípicamente en 4 ambientes. La interacción $G \times A$ es la componente de varianza más importante. La primera estrategia fue estimar modelos GWAS por ambiente, considerando la estructura genética a través de la matriz de pedigree y, alternativamente, mediante la similitud molecular (EMMA). La segunda estrategia fue ajustar un modelo GWAS para el total de datos donde se incorpora $G \times A$, usando la correlación de efectos genómicos entre ambientes para contemplar la interacción; nuevamente la matriz de relaciones aditivas fue calculada a partir del pedigree y de la matriz de marcadores moleculares. El Mejor Predictor Lineal Insesgado (BLUP) de los efectos de G en cada A fue derivado desde cada modelo. Se correlacionaron los BLUPs obtenidos bajo distintas estrategias de modelación con los obtenidos en el modelo de referencia. Con la primera estrategia implementada, para ambas formas de estimar la matriz de relaciones aditivas, se obtuvieron altas correlaciones de

la performance estimada de cada G en cada A con la performance observada a partir del modelo de referencia.

Palabras claves: componentes de varia; blup; interacción g×a; gwas

PROPUESTA DE DISEÑO DE UN SISTEMA PARA EL SEGUIMIENTO ACADÉMICO DE ESTUDIANTES Y DOCENTES DE LA CARRERA DE ING. EN SISTEMAS DE INFORMACIÓN - UTN FRC

***CECILIA SAVI, ANDREA FABIANA RIGHETTI, ANA MARÍA STRUB, CLARISA
LILIANA STEFANICH y IRENE ESTHER ROMOLI***

ceciliasavi@hotmail.com

La acreditación es un proceso por medio del cual un programa o institución educativa brinda información sobre sus operaciones y logros a un organismo externo que evalúa y juzga. La evaluación y la acreditación son procesos relacionados cuya práctica se entrecruzan, ya que se acredita conforme y como consecuencia de un proceso de evaluación y seguimiento. La acreditación no sólo provee un diagnóstico que conduce a la acción por parte de la propia institución, sino que constituye una constancia de credibilidad por parte de la sociedad demandante de los servicios educativos. Estos servicios educativos implican, conceptos relativos, intangibles y muchas veces subjetivos, que no permiten una evaluación concreta y/o absoluta, lo que obliga al diseño de mecanismos de control diferentes. Una opción posible es la creación de indicadores estadísticos que permiten relacionar funcionamiento, recursos y resultados respecto a actividades, procesos, unidades organizacionales y otros componentes de una institución. La evaluación de instituciones y de carreras de educación superior, en sí misma, constituye un valioso instrumento que proporciona elementos de juicio para analizar a fondo los procesos educativos, generando información para promover y asegurar la mayor calidad, productividad y pertinencia de las acciones y resultados. En este contexto y como parte del diseño de un Sistema de Seguimiento Académico de estudiantes y docentes de la carrera Ingeniería en Sistemas de Información UTN-FRC, se presenta una propuesta de indicadores estadísticos específicos. Para ello en una primera etapa se analizan, definen y conceptualizan dimensiones de análisis de la calidad educativa, las variables interviniéntes y los indicadores, precisando nombre, definición, explicación, fuentes de obtención de la información de base, fórmula de cálculo, y elementos que se consideren necesarios para la interpretación de los resultados. Se propone además, construir una aplicación informática que permita capturar y procesar datos del Sistema Académico de la FRC e ingresar los que se generen a través de instrumentos de recolección específicos. El diseño del software prevé la presentación de información mediante tablas, cuadros comparativos, gráficos estadísticos, tableros de comando y archivos exportables, para distintos usuarios de la gestión académica. La construcción de la herramienta de software para el tratamiento de datos masivos pretende contribuir a la necesidad de manipular enormes cantidades de datos y también a la creación de modelos predictivos para ser utilizados específicamente en la carrera propuesta.

Palabras claves: indicadores estadíst; seguimiento académic; seguimiento académic; dimensiones de anál

**PROPUESTA PARA LA IMPLEMENTACIÓN DEL
SOFTWARE R EN LAS CARRERAS DE CONTADOR
PÚBLICO Y LICENCIATURA EN ADMINISTRACIÓN DE
LA FACULTAD DE CIENCIAS ECONÓMICAS Y
ESTADÍSTICA DE LA UNIVERSIDAD NACIONAL DE
ROSARIO**

**MARÍA EVANGELINA ÁLVAREZ, JUAN JOSÉ CÁMPORA y NORMA BEATRIZ
MEROI**

FACULTAD DE CIENCIAS ECONÓMICAS Y ESTADÍSTICA, UNIVERSIDAD NACIONAL DE ROSARIO.

normamerol@outlook.com

La sociedad actual, llamada de la información, demanda cambios en los sistemas educativos de forma que éstos se tornen más flexibles y accesibles, menos costosos y a los que han de poderse incorporar los ciudadanos en cualquier momento de su vida. Nuestras instituciones de formación superior, para responder a estos desafíos, deben revisar sus referentes actuales y promover experiencias innovadoras en los procesos de enseñanza-aprendizaje apoyados en las Tecnologías de la Información y la Comunicación (TIC). Y, contra lo que estamos acostumbrados a ver, el énfasis debe hacerse en la docencia, en los cambios de estrategias didácticas de los profesores, en los sistemas de comunicación y distribución de los materiales de aprendizaje, en lugar de enfatizar la disponibilidad y las potencialidades de las tecnologías. Las simulaciones interactivas, los recursos educativos digitales y abiertos, los instrumentos de recolección y análisis de datos son algunos de los muchos recursos que permiten a los docentes ofrecer a sus estudiantes posibilidades, antes inimaginables, para asimilar conceptos. En la asignatura Métodos Estadísticos y Estadística para Administradores de la Facultad de Ciencias Económicas y Estadística los docentes nos estamos replanteando nuestro quehacer en el aula y evaluando la factibilidad de implementar el uso de nuevas tecnologías mediante el programa R Commander en el dictado de las mismas. En una primera etapa del proyecto de investigación llevamos a cabo un relevamiento bibliográfico para conocer esta herramienta y las ventajas y desventajas en la implementación de este software y su uso en algunas facultades de nuestro país. Además proponemos un primer material de trabajo para confeccionar cuadros y gráficos estadísticos y realizar análisis descriptivo univariado para variables cuantitativas. Uno de los objetivos de nuestro trabajo de investigación es que los alumnos conozcan la herramienta y logren optimizar el tiempo; enfatizando más en las interpretaciones que en la realización de cálculos estadísticos extensos.

Palabras claves: r commander; métodos estadísticos; licenciatura en admi; ntic

**RENDIMIENTO ACADÉMICO: ANÁLISIS DE
TRAYECTORIAS EN MATEMÁTICA Y BIOESTADÍSTICA**

**MARCELA FERNICOLA¹, MYRIAM NÚÑEZ¹, CHRISTIANE PONTEVILLE¹ y
FERNANDA RUSSO²**

¹FACULTAD DE FARMACIA Y BIOQUÍMICA. UBA. ²FACULTAD DE FARMACIA Y BIOQUÍMICA. UBA.

chponteville@gmail.com

El objetivo del presente trabajo es analizar la evolución del rendimiento académico de los estudiantes de las carreras de Farmacia y Bioquímica de la Universidad de Buenos Aires, inscriptos a las asignaturas Matemática y Bioestadística en ambos cuatrimestres del año 2015. Para este análisis

se recabaron datos sobre el resultado de la prueba diagnóstica inicial que se realiza el primer día de clase, confeccionada para evaluar los saberes previos necesarios para la cursada de la asignatura correspondiente. Se efectuó una partición de los resultados en tres grupos, considerando la proporción de ejercicios bien resueltos. Asimismo, se relevaron datos relativos a la trayectoria de cursada: calificaciones parciales y resultado en instancias de recuperación, condición final de cursada y calificación en exámenes finales. Se observó que en Matemática los niveles de promoción (nota promedio de parciales 7 o más) se mantuvieron prácticamente dentro de los mismos márgenes en ambos cuatrimestres: 10%, 36% y 45% en el primero, frente al 13%, 33% y 52% respectivamente en el segundo, sobre el total de aprobados por grupo, mientras que en el primer cuatrimestre alcanzaron la regularidad (promocionaron o aprobaron los trabajos prácticos) un 47%, 69% y 81% de los alumnos, frente al 36%, 69% y 76% respectivamente del segundo. Por su parte, en Bioestadística, promocionaron el 34%, 58% y 66% del total de regularizados de cada grupo en el primer cuatrimestre, mientras que el 40%, 36% y 59% lo hizo en el segundo. Los niveles de regularización fueron del orden del 72%, 80% y 84% en el primer cuatrimestre, contra un 58%, 65% y 78% en el segundo. Se concluyó que en ambas asignaturas la evaluación inicial proporciona una herramienta de suma utilidad para pronosticar el rendimiento académico posterior, por ejemplo, se observó que en aquellos casos con pobre desempeño inicial, la cursada continúa presentándose con dificultad. La falta de conocimientos adquiridos en los niveles anteriores de formación permite pensar en el uso de algunas estrategias pedagógicas diferentes para este grupo.

Palabras claves: bioestadística; matemática; rendimiento académico; prueba diagnóstica

REVISIÓN SISTEMÁTICA DE LOS EFECTOS PRODUCIDOS POR CAMBIOS EN LA DIETA ALIMENTARIA SOBRE LA PRESIÓN ARTERIAL

LUCÍA CARABALLO, AYLÉN AVILA, DARÍO WEITZ, FERNANDO DÍAZ PACÍFICO, MARTA MARZI, LAURA PISKULIC y MARÍA BELEN ALLASIA

FACULTAD DE CIENCIAS BIOQUÍMICAS Y FARMACÉUTICAS, UNIVERSIDAD NACIONAL DE ROSARIO.

mallasia@fbiogyf.unr.edu.ar

La hipertensión arterial es el principal factor de riesgo para la enfermedad cardiovascular. La adopción de una dieta con bajo contenido de sodio, grasas e hidratos de carbono y abundante en fibras naturales ha sido descripta como una estrategia para ayudar a prevenir o controlar la hipertensión. El objetivo del presente trabajo fue recolectar y sistematizar la evidencia acerca de los efectos que cambios en la dieta producen sobre la presión arterial. Se realizó una Revisión Sistemática (RS) de estudios publicados hasta abril de 2016 a partir de búsqueda electrónica en Medline/PubMed, SciELO y LILACS. Siguiendo los lineamientos del Manual Cochrane de Revisiones Sistemáticas de Intervenciones, se incluyeron ensayos controlados aleatorizados (ECAs) que evaluaron los efectos de modificaciones en la dieta sobre la presión arterial, comparando una dieta intervención (DI), rica en frutas y verduras y reducida en grasas, carnes rojas y azúcares, con la ingesta habitual de los pacientes (dieta control, DC). Criterios de elegibilidad: población adulta sin restricción de edad, sexo, etnia; presión arterial mayor a 120/80 mmHg sin medicación antihipertensiva; sometidos a tratamientos dietarios entre 2 semanas y 6 meses de duración. Se identificaron 1.077 estudios, de los cuales 10 satisfacían los criterios de elegibilidad y fueron incluidos en la RS, contribuyendo con un total de 1.541 participantes (DC: n=801, DI: n=783). El efecto, medido como la diferencia de la reducción media de la presión arterial entre el grupo intervención y el grupo control, fue significativo en 9 de los 10 estudios individuales tanto para la presión sistólica (PS) como para la presión diastólica (PD). En 8 de los 10 estudios, la magnitud de dicho efecto fue mayor para la PS que para la PD, tanto en pacientes prehipertensos como hipertensos. La RS reveló que una dieta rica en frutas, verduras y productos lácteos bajos en grasa, con granos enteros y carnes ma-

gras, pero reducida en dulces y carnes rojas tiene significativos efectos favorables sobre la presión sistólica y diastólica como primera línea de tratamiento antihipertensivo. Sin embargo, la heterogeneidad estadística entre los resultados publicados en los ECAs, la imprecisión de la información publicada en algunos artículos y la imposibilidad de contactar a los autores para obtener los datos originales, impidió realizar un Metanálisis a fin de cuantificar el efecto global promedio.

Palabras claves: revisión sistemática; hipertensión arteria; dieta alimentaria

SELECCIÓN DE VARIABLES PARA LA MODELACIÓN DE LA DISTRIBUCIÓN POTENCIAL DE ECHINOPSIS ATACAMENSIS UTILIZANDO MAXENT

SILVIA SÜHRING¹, JESÚS SAJAMA², SERGIO FONTEÑEZ³, ANDREA BARRIONUEVO⁴, PABLO GOROSTIAGUE⁵ y PABLO ORTEGA-BAES⁵

¹LABORATORIO DE INVESTIGACIONES BOTÁNICAS (LABIBO), FAC. DE CS. NATURALES, U.N.DE SALTA Y CÁTEDRA DE ESTADÍSTICA Y DISEÑO EXPERIMENTAL, FAC. DE CS. NATURALES, U.N.DE SALTA. ²LABORATORIO DE INVESTIGACIONES BOTÁNICAS (LABIBO), FAC. DE CS. NATURALES, U.N.DE SALTA Y CÁTEDRA DE CÁLCULO ESTADÍSTICO, SEDE REGIONAL ORÁN, FAC. DE CS. NATURALES, U.N.DE SALTA.

³CÁTEDRA DE ESTADÍSTICA Y DISEÑO EXPERIMENTAL, FAC. DE CS. NATURALES, U.N.DE SALTA. ⁴LABORATORIO DE INVESTIGACIONES BOTÁNICAS (LABIBO), FAC. DE CS. NATURALES, U.N.DE SALTA; CÁTEDRA DE ESTADÍSTICA Y DISEÑO EXPERIMENTAL, FAC. DE CS. NATURALES, U.N.DE SALTA Y CÁTEDRA DE CÁLCULO ESTADÍSTICO. ⁵LABORATORIO DE INVESTIGACIONES BOTÁNICAS (LABIBO), FAC. DE CS. NATURALES, U.N.DE SALTA.

ssuhring@gmail.com

Los modelos de distribución potencial permiten predecir las áreas de presencia de una especie con base en variables ambientales. MaxEnt es uno de los modelos más precisos, sólo requiere puntos de registro georeferenciados de la especie y coberturas geográficas de variables ambientales que podrían limitar su supervivencia. El programa asigna a cada celda de un mapa un valor continuo de probabilidad de aptitud de hábitat, luego se determina un valor umbral de corte para reclasificar esta probabilidad y obtener mapas de presencia/ausencia (mapas de distribución potencial ajustada, MDPA). En muchos casos la selección de variables depende de la disponibilidad de su representación geográfica. El número de variables incluidas en los modelos afecta las predicciones finales, en general el área predicha disminuye si se incluyen más variables. Por otro lado, algunas variables pueden resultar redundantes, dada la asociación que existe entre sus valores. En este trabajo se evaluó el efecto de la cantidad de variables incluidas en la modelación y de la colinealidad existente entre ellas, sobre los MDPA obtenidos para *Echinopsis atacamensis*, una cactácea columnar nativa de Argentina. Se comparó el rango geográfico, latitudinal, longitudinal y altitudinal de los MDPA, y se evaluó la similaridad de nicho utilizando el estadístico D de Warren. Se utilizaron las 19 variables climáticas tomadas de WORLDCLIM, referidas principalmente al comportamiento de la temperatura y las precipitaciones, y la altitud. Con estas variables se realizó un ACP para obtener componentes ortogonales. Se generaron siete modelos utilizando distintos conjuntos de variables: 1) todas; 2) excluyendo las correlacionadas con más variables; 3) excluyendo las correlacionadas con las que más contribuyeron al modelo 1; 4) las de mayor peso en el CP1; 5) las de mayor peso en CP1 y CP2; 6) las CP 1 a 4 (autovalores λ); 7) las CP 1 a 7 (λ ; 99% de varianza explicada). A cada modelo se le aplicó el umbral de corte tal que incluya el 90 % de los registros de presencia (percentil 10). Los MDPA correspondientes a los modelos 3 y 4 tuvieron mayor área (1,5 veces respecto del MDPA 1), y mayor rango altitudinal, latitudinal y longitudinal (1,1 a 1,46 veces respecto del MDPA 1). El área fue menor para los modelos que incluyeron mayor cantidad de variables. Todos los MDPA tuvieron una alta superposición de nicho (D λ 0,87), resultando mayor entre los MDPA 1, 2 y 5, los que además presentaron menor rango latitudinal y longitudinal.

Palabras claves: análisis de componen; mapas de distribuciόn; cactaceae; colinealidad

TRASTORNOS DEL ESPECTRO AUTISTA ESTIMACIÓN DE SU PREVALENCIA A EDADES TEMPRANAS EN LA CIUDAD DE SANTA FE, COSTOS DE TRATAMIENTO Y EFECTOS DE LA ACTIVIDAD FÍSICA SOBRE HABILIDADES MOTORAS GRUESAS

LILIANA ESTER CONTINI¹, FRANCISCO ASTORINO², DIEGO CARLOS MANNI³, GABRIEL FESSIA² y ELENA FERNANDEZ DE CARRERA²

¹FACULTAD DE BIOQUÍMICA Y CIENCIAS BIOLÓGICAS - UNL. ²FACULTAD DE CIENCIAS MÉDICAS - UNL. ³FACULTAD DE BIOQUÍMICA Y CIENCIAS BIOLÓGICAS - FACULTAD DE CIENCIAS MÉDICAS - UNL.

lcontini@fbcn.unl.edu.ar

El autismo es un complejo trastorno del desarrollo psicomotor que comienza antes de los tres años de edad y persiste durante toda la vida del individuo. La importancia de la detección precoz de los niños con trastornos del espectro autista (TEA) permite la intervención oportuna en etapas iniciales de la vida impactando positivamente en la calidad de vida del paciente y las de sus familias. No se conoce con exactitud la prevalencia de TEA. Según la OMS en las Américas es de 65,5/10.000 (34-90/10.000). Esta estimación representa una cifra media, pues la prevalencia observada difiere considerablemente entre las distintas investigaciones, si bien, en estudios bien controlados se han registrado cifras notablemente mayores. En Argentina no se tienen datos al respecto. Esta información es primordial para establecer las bases de una propuesta de Salud Pública que permita lograr diagnósticos precoces e intervenciones oportunas, como lo expresan las leyes Nacional y de la Provincia de Santa Fe de Autismo. Por esta razón, un grupo de investigadores multidisciplinario de la Universidad Nacional del Litoral, realizó una investigación con el fin de determinar la prevalencia de TEA en niños pequeños en la ciudad de Santa Fe, así como también se propuso evaluar los beneficios de la actividad física en niños con TEA y los costos directos e indirectos destinados a diagnóstico, tratamiento, educación. El objetivo de este trabajo es presentar los hallazgos de esta investigación. Prevalencia: se realizó un muestreo estratificado con afijación proporcional en la ciudad. Se evaluaron 506 niños de 18 a 36 meses de edad resultando una prevalencia temprana estimada de 1 cada 128, valor que se encuentra dentro del rango dado por la OMS para las Américas. Costos: Si bien este ítem es muy difícil de determinar debido a la multiplicidad de terapias, sistemas de salud y de educación, se realizó una encuesta, de respuesta voluntaria, a padres de niños con TEA de provincias de Argentina. Sólo fueron respondidas 28 de más de 50 entregadas. De ellas se pudo concluir que el rango de los gastos por niño por mes es: 1809,44–24427,44. Información que no puede compararse, no se conocen estudios similares en Argentina. Actividad física: De la evaluación de la experiencia realizada en adolescentes con TEA, se encontró que en todos los casos hubo aumento de las habilidades motoras gruesas que variaron entre 40,0%-85,7% de mejoría respecto de las de habilidades iniciales.

Palabras claves: tea; prevalencia; epidemiología

VALIDACIÓN DE UN INSTRUMENTO DISEÑADO EN UNA CÁTEDRA DE CIRUGÍA VETERINARIA PARA INNOVAR EN LA TOMA DE EXÁMENES ORALES

ELENA TERESITA FERNÁNDEZ DE CARRERA¹ y JORGE FIORENTINI²

¹FACULTAD DE CIENCIAS MÉDICAS - UNL. ²FACULTAD DE CIENCIAS VETERINARIAS, UNR.

elenacarrera2@gmail.com

En las cátedras de Cirugía I y II de la Facultad de Ciencias Veterinarias (FCV) de la Universidad Nacional de Rosario, se planteó la necesidad de modificar la metodología de la evaluación final de ambas materias que fundamentalmente eran exámenes orales. Estos se realizaban en la forma tradicional siguiendo las pautas puestas en práctica desde el comienzo. Pero surgieron los interrogantes: ¿el método para evaluar estos exámenes finales era el mismo en todos los llamados y para todos los alumnos o dependen de la subjetividad? ¿Todos los docentes asignan la misma jerarquía a todos los ítems a evaluar? Para ello se realizó una investigación de la temática señalada que además propondría la confección de una planilla que guiara a los docentes en la forma de evaluar para homogeneizar criterios y minimizar el impacto de la subjetividad. Objetivo: Evaluar la validez de la planilla diseñada como instrumento de evaluación para guiar y ordenar el desarrollo del examen final oral de las asignaturas Cirugía I y II. Metodología y resultados : Se realizó un análisis de la concordancia de las calificaciones de 100 exámenes tomados por dos evaluadores distintos con empleo de la planilla diseñada al efecto mediante la aplicación del análisis de regresión entre las notas de ambos evaluadores, obteniéndose un coeficiente de correlación lineal de Pearson de 0,929 ($p < 0,001$). A continuación se analizaron los indicadores empleados en la planilla para identificar aquellos que presentaban mayor diferencia. Interesa también investigar la “concordancia” de las notas puestas por cada evaluador. Por último se los analiza como variable continua y resulta que no hay acuerdo en las notas de sólo 5 alumnos dado que la media de la diferencia de notas entre ambos evaluadores es 0,9 con un intervalo de confianza del 95%, según proponen Bland y Altam de (-10,8; 12,7). Conclusión: El uso de la planilla es concordante entre ambos evaluadores pero existen problemas en alguno de los indicadores que convendría revisar. Los más frecuentes son el indicador 2, el 3 y el 4. Se sugiere unificar la puntuación de los indicadores para facilitar el uso de la planilla.

Palabras claves: exámenes orales; cirugía veterinaria; validación de instru

ÍNDICE DE AUTORES

A

- ABBIATI, N 109
 ABBIATI, NN 105
 ABRIL, JC 49
 ABRIL, MDLM 49
 ACHCAR, JA 19
 ACIÓN, L 59
 AGUZZI, V 142
 AHUMADA, MI 65
 ALESSO, C 79
 ALLASIA, MB 72, 90, 120, 140, 156
 ALTEZ DE CASTRO, Y 113
 ALVAREZ, EE 30
 ALVAREZ, L 93
 ALVAREZ, ME 56, 62, 155
 ALVAREZ, R 34, 113
 ALVAREZ-CASTRO, I 30
 ANCHORENA, J 44
 ANDINO, N 127
 ANDREOZZI, L 146
 ANGELINI, J 20, 92
 ANTILLE, D 79
 ARANDA, E 124
 ARAYA ALPÍZAR, C 33
 ARIAS, V 65, 143
 ARNESI, N 86, 133
 ARROYO, L 92
 ARRUFAT, JL 49
 ARTUSO, C 28
 ASTORINO, F 158
 AVILA, A 156
 AVILA, OB 54, 72, 113, 146

B

- BÓ, G 122
 BÓBEDA, G 134
 BALPARDA, L 78, 142
 BALZARINI, M 74, 115, 122, 125, 153
 BANGDIWALA, S 53
 BAQUERO, R 115
 BARBARÁ, G 43
 BARBONA, I 95, 128
 BARENGO, PB 89
 BARONIO, A 99
 BARRIONUEVO, A 157
 BASSOLS, G 84
 BATTISON, N 76
 BATTISTON, N 79
 BAYÓN, C 94
 BELTRÁN, C 95

- BENAVIDES, E 118
 BERRETTA, G 107
 BERTA, EE 146
 BERTOLA COMPAGNUCCI, A 96
 BLACONÁ, MT 146
 BLANCO, G 107
 BOCA, T 80
 BOGGIO, G 20, 133
 BONFILI, O 69, 70
 BORRA, VL 46
 BOSANO, J 48
 BOTTAI, H 72, 140
 BRAMARDI, S 34, 108, 134, 141, 149
 BRANCO, MD 90
 BRAUN, S 115
 BRUNO, C 91, 153
 BRUZZONE, O 135
 BURGUETE EGUREN, V 113
 BUSSI, J 106
 BUZZI, SM 49, 50

C

- CÁCERES BAUER, RD 18
 CÁMPORA, JJ 155
 CÉSPEDES SOLÍS, ADLA 31
 CÓRDOBA, M 74
 CABODEVILA, VG 97
 CABRERA, GP 38
 CACCHIARELLI, P 85
 CACHEIRO, L 122
 CALVIÑO, A 94
 CALZADA, J 118
 CAMBIASO, V 85
 CAMIZ, S 35
 CAMUSSI, M 142
 CANALIS, P 120
 CANCINOS, M 120
 CARABALLO, L 156
 CARCEDO, D 77
 CARNEVALI, GH 62
 CARO, NP 65, 99, 114, 117
 CARRIZO, M 79
 CASANOVES, F 150
 CASPARI, MT 28
 CASPERRI, MT 48
 CASTELLANA, N 121
 CASTIGLIONI, L 141
 CASTILLO SOSA, J 103
 CASTILLO, M 118
 CASTRO KURISS, C 21

| | |
|------------------------|---------|
| CASTRO, MJ | 83 |
| CATALANO, AM | 44 |
| CATALANO, M | 121 |
| CAUMONT, MP | 151 |
| CERINO, RM | 148 |
| CERVIGNI, GDL | 92 |
| CHACÓN, C | 121 |
| CHAILA, S | 126 |
| CHAPARRO, M | 24 |
| CHAVERO, M | 142 |
| CHAVES, L | 118 |
| CHIAPELLA, LC | 87 |
| CHUMPITAZ, A | 142 |
| CIAPPONI, A | 68 |
| CIARLO, E | 124 |
| CICCIOLI, P | 106 |
| CIPRIOTTI, P | 79, 80 |
| CLAUSSE, A | 109 |
| CLOSAS, AH | 101 |
| CONDA, C | 131 |
| CONTINI, LE | 72, 158 |
| CONTRERAS, D | 103 |
| COOK, D | 36 |
| CORALLO, L | 78 |
| CORDOBA, M | 73 |
| CORRÍAS, C | 151 |
| COSTA TÁRTARA, S | 34 |
| COTTET, P | 86 |
| COVAS, MDC | 81 |
| CRAVERO, M | 60 |
| CROMBET, T | 128 |
| CUCCIARELLI, L | 140 |
| CUESTA, C | 128 |
| CUESTA, CB | 68 |
| CUETO, G | 115 |
| CURETTI, M | 141 |
| CURRETTI, M | 149 |
| CURTI, R | 34 |

D

| | |
|-------------------------|--------|
| D'ERCOLE, FN | 99 |
| D'URSO VILLAR, MA | 53 |
| DÍAZ PACÍFICO, F | 156 |
| DÍAZ, DE | 58 |
| DÍAZ, M | 41 |
| DÍAZ, MDP | 51 |
| D'IORIO, S | 54 |
| D'IORIO, S | 113 |
| D'AMELIO, A | 151 |
| DA SILVA, N | 36, 59 |
| DAMIANO, L | 32 |
| DAMILANO, G | 55 |
| DAORDEN, ME | 92 |

| | |
|-----------------------------|---------|
| DAURELIO, L | 89, 139 |
| DE FRANCESCHI, M | 110 |
| DE GIUSTO, G | 72 |
| DE LOS RÍOS, C | 41, 127 |
| DE OLIVEIRA PERES, MV | 19 |
| DE OLIVEIRA, R | 149 |
| DEL BRÍO, D | 108 |
| DEL MEDICO, AP | 97, 104 |
| DELFINO, H | 44, 130 |
| DELFINO, VM | 119 |
| DEMETRIO, PM | 112 |
| DEPETRIS, J | 97 |
| DI MASSO, R | 123 |
| DI RIENZO, JA | 73, 150 |
| DIAZ, M | 143 |
| DIGONZELLI,, PA | 138 |
| DIMARCO, J | 142 |
| DIMATTIA, J | 124 |
| DRUNDAY, F | 94 |
| DUARTE, D | 36 |
| DUARTE, M | 57 |
| DUCTARY, S | 118 |
| DURÁN, L | 137 |
| DURET, G | 56 |

E

| | |
|--------------------|-----|
| EASDALE, M | 135 |
| ECHALAR, SR | 82 |
| ELISEI, N | 93 |
| ELORZA, E | 86 |
| ESCANES, V | 130 |
| ESPARIZ, M | 89 |
| EUN-KYUNG, L | 36 |

F

| | |
|---------------------------------|----------|
| FAGNOLA, B | 97 |
| FARACE, ML | 71 |
| FERNÁNDEZ DE CARRERA, ET | 158 |
| FERNÁNDEZ, EDV | 38 |
| FERNÁNDEZ, EN | 105 |
| FERNÁNDEZ, JI | 71 |
| FERNÍCOLA, M | 94 |
| FERNANDEZ DE CARRERA, E | 158 |
| FERNANDEZ DE ULLIVARRI, J | 138 |
| FERNANDEZ, D | 141, 149 |
| FERNENDEZ, M | 41 |
| FERNICOLA, M | 155 |
| FERRERI, NM | 56, 62 |
| FESSIA, G | 158 |
| FILIPPINI, OS | 119, 130 |
| FILIPPINI, S | 106, 110 |
| FIORENTINI, J | 158 |
| FONTEÑEZ, S | 157 |

| | |
|------------------------|----|
| FONTEÑEZ, SL | 82 |
| FRERY, A | 60 |
| FRETES CHÁVEZ, A | 78 |
| FRIDRIJ, GA | 28 |

G

| | |
|--------------------------|--------------|
| GÓMEZ, M | 41 |
| GARCÍA, F | 76, 91 |
| GARCÍA, G | 41, 145 |
| GARCÍA, LA | 94 |
| GARCÍA, MDC | 121 |
| GARCIA, LA | 81 |
| GARCIA, MB | 138 |
| GARCIA, MDC | 123 |
| GARIBOTTI, G | 75 |
| GAZZERA, G | 141 |
| GEVAERT, O | 144 |
| GIANNINI KURINA, F | 125 |
| GIARDINI, P | 142 |
| GIMÉNEZ, G | 66, 108, 152 |
| GIMÉNEZ, L | 134 |
| GIMÉNEZ, P | 40 |
| GIORGINI, DA | 106 |
| GIRIMONTE, P | 93 |
| GIULIANO, M | 107 |
| GONZALEZ, C | 83 |
| GONZALEZ, CP | 105 |
| GONZALEZ, LA | 67 |
| GONZALEZ, MV | 148 |
| GOROSTIAGUE, P | 157 |
| GOYENECHE, JJ | 25, 47 |
| GUARDIOLA, M | 99, 114 |
| GUARRACINO, L | 40 |
| GUEDES, G | 36 |
| GUERRERO, F | 102 |
| GUILLON, MDLP | 81 |

H

| | |
|----------------------|-----|
| HÚMPOLA, MV | 72 |
| HANG, S | 125 |
| HARVEY, GB | 68 |
| HEREDIA, M | 39 |
| HERNÁNDEZ, A | 94 |
| HERNÁNDEZ, L | 43 |
| HERNANDEZ, L | 23 |
| HERRERA, MDC | 82 |
| HIDALGO, M | 83 |
| HIRSCHLER, V | 83 |
| HOTTA, LK | 50 |
| HUAYLLA, C | 75 |
| HURTADO A., LA | 118 |

I

| | |
|-------------------|----|
| IGLESIAS, J | 91 |
|-------------------|----|

| | |
|----------------------|---------|
| IGLESIAS, M | 118 |
| IMHOFF, S | 79 |
| INCOCCIATI, JP | 53 |
| ISERN, G | 96, 128 |
| IVANCOVICH, JJ | 72, 140 |
| IZAGUIRRE, A | 27 |

J

| | |
|-------------------------|-----|
| JIMÉNEZ ALFARO, E | 136 |
| JOEKES, S | 17 |
| JUÁREZ, MA | 97 |
| JUAN, RD | 73 |
| JURY, S | 93 |

K

| | |
|----------------------|-----|
| KATZ, RD | 56 |
| KELMANSKY, DM | 42 |
| KUCUKBEYAZ, DR | 53 |
| KUCUKBEYAZ, R | 141 |

L

| | |
|----------------------------|--------------|
| LÓPEZ ARMENGOL, MF | 152 |
| LÓPEZ RÍOS, VI | 24 |
| LÓPEZ, D | 78, 142 |
| LÓPEZ, N | 100 |
| LÓPEZ, R | 125 |
| LABAT, SD | 58 |
| LACAZE, MV | 116 |
| LAFFITTE, L | 138 |
| LALLEMENT, M | 75 |
| LAMFRI, M | 78 |
| LAQUIDARA, AP | 93 |
| LARGO, R | 115 |
| LASPINA, M | 97 |
| LASTIRI, JM | 98 |
| LAVALLE, A | 97, 104, 138 |
| LAZARTE, VF | 28 |
| LEIVA, V | 21 |
| LEJARRAGA, H | 42 |
| LENCINA, V | 39 |
| LENCINA, VB | 46 |
| LLANCALAHUEN, M | 70 |
| LOGIOCO, C | 83 |
| LOPEZ ARMENGOL, MF | 66 |
| LORENZO-LUACES, P | 128 |
| LOVATTO, M | 52 |
| LUCERO BRINGAS, MDLA | 117 |
| LUNA, LI | 65, 91, 143 |
| LUPÍN, B | 65 |

M

| | |
|---------------------|--------|
| MACCALLINI, G | 83 |
| MAGLIONE, DS | 69, 70 |
| MALDONADO, E | 107 |

- MAMPRIN, ME 87
 MANNI, DC 158
 MANTARAS, R 124
 MANZOTTI, M 76, 79
 MARÍ, G 43
 MARCO, LA 38
 MARFETÁN MOLINA, D 71, 100
 MARIÉ, D 24
 MARIANO, SM 151
 MARQUEZ, V 115
 MARTÍN, MC 19
 MARTÍNEZ, RD 105
 MARTINEZ, J 60
 MARZI, M 156
 MASAUTIS, A 42
 MASCI, ME 48
 MASERA, P 143
 MASOLA, M 79
 MASSA, F 21, 34
 MATTALIA, M 22
 MAYANS, L 138
 MAZZA, S 134
 MEDINA, O 109
 MEINARDI, BL 100
 MEJÍA, R 115
 MELO, O 116
 MELO, SP 134
 MENA, J 39
 MENDEZ, F 125
 MEROI, N 128
 MEROI, NB 155
 MESA, L 102
 METELLI, MA 28
 MICHELETTI, L 124
 MIGNONI, CA 46
 MILÁN, C 76, 79
 MILANI, H 149
 MILLANES, I 120
 MITAS, G 43
 MODINI, L 102
 MOLINA, J 138
 MOLINARI, C 83
 MONETA PIZARRO, AM 97
 MONTENEGRO BRUSOTTI, J 84
 MONTERO, L 97
 MONTES ROJAS, G 27
 MONTES, A 128
 MORALES, D 67
 MORALES, E 41
 MORALES, JM 88
 MORENO, L 25
 MORO, A 142
 MUÑIZ SAAVEDRA, J 138

- MUÑOZ MORALES, D 67
N
 NÚÑEZ, M 155
 NAIDICZ, PL 28
 NATAL, M 24
 NENINGER, E 128
 NIEMI, J 30
 NOLASCO, M 77
 NUÑEZ, M 83, 84, 94, 98
 NUNES, A 149
 NUNES, F 42
O
 OÑA, A 37
 OCAMPO, S 138, 141, 149
 OCHOA, MDC 126
 ODRIozola, JG 101
 OJEDA, S 26
 OJEDA, SM 50
 ORTEGA-BAES, P 157
 ORTIZ, NDV 138, 147
 ORTIZ, P 99
 OZÁN, NS 41
P
 PÉREZ, A 115
 PACCIORETTI, P 67, 74
 PACCIORETTI, PA 150
 PADRÓ, OB 51
 PAGURA, J 23
 PAGURA, JA 46, 145
 PALACIOS, L 106, 110
 PALLOTTA, G 98
 PASSERINI, MS 83
 PAZ, S 109
 PEÑA MALAVERA, A 38
 PEREA, M 118
 PEREIRA DA COSTA, J 85
 PEREIRA, W 36
 PERES, MVO 149
 PERETTI, ME 142
 PEREZ, N 135
 PEZZOTTO, SM 96
 PIA, A 73
 PICARDI, DL 136
 PICCARDI, M 122
 PINEDA RIOS, WD 29
 PISANI, MV 86
 PISKULIC, L 72, 139, 140, 156
 PISTONESI, S 60
 PIZARRO, MD 89
 POK, C 57
 PONTEVILLE, C 155

| | |
|--------------------------|------------------------|
| PORCASI, X | 78 |
| PRARIZZI, M | 79 |
| PRATTA, GR | 85, 97, 104 |
| PREVIDELLI, I | 149 |
| PRIETO ANGUEIRA, S | 126 |
| PRONELLO, FA | 148 |
| PRUNELLO, M | 72, 124, 139, 140, 144 |
| Q | |
| QUAGLINO, MB | 35, 90, 92, 120 |
| R | |
| RAÑA, E | 126 |
| RACCA, L | 72, 140 |
| RAMPOLDI, A | 125 |
| RAPELLI, C | 121, 123 |
| REALPE CAMPAÑA, JD | 102 |
| REDONDO, Y | 61 |
| REFOJO, D | 109 |
| RIAÑO, ME | 26 |
| RIBEIRO, M | 149 |
| RIBEIRO, R | 36 |
| RICCI, L | 24, 26 |
| RIDDICK, ML | 30 |
| RIGG, C | 118 |
| RIGHETTI, A | 17 |
| RIGHETTI, AF | 154 |
| RIGO, D | 55 |
| ROBERTS, KN | 68 |
| RODRÍGUEZ, C | 128 |
| RODRÍGUEZ, GR | 85 |
| RODRIGUEZ VELEZ, A | 122 |
| RODRIGUEZ, MDL | 83 |
| ROJAS, C | 115 |
| ROJO, MP | 111 |
| ROMERO, E | 138 |
| ROMERO, MDC | 109 |
| ROMOLI, IE | 154 |
| ROSA, EA | 45 |
| ROSA, MA | 22 |
| ROSATI, G | 44 |
| ROSSI, L | 151 |
| ROVEGNO, MS | 105 |
| RUBIO, N | 66, 108, 152 |
| RUEDA CALDERÓN, MA | 153 |
| RUIZ SUAREZ, S | 88 |
| RUIZ, L | 133 |
| RUSSO, F | 98, 155 |

| | |
|-------------------|-----|
| S | |
| SÜHRING, S | 157 |
| SÁENZ, JL | 70 |
| SÁNCHEZ, JP | 105 |
| SAINO, M | 103 |

| | |
|----------------------------|--------------|
| SAJAMA, J | 157 |
| SALCEDO PALACIOS, TE | 29 |
| SALOMÓN, RCJ | 134 |
| SALVATIERRA, P | 76, 79 |
| SAN MARTIN, M | 23 |
| SANDOVAL, M | 70 |
| SANTELLÁN, S | 60 |
| SARAVIA, J | 130 |
| SARMIENTO, PV | 22 |
| SAVI, C | 154 |
| SCAVINO, M | 25 |
| SCHLAPS, E | 60 |
| SEMPRINI, DG | 22 |
| SERRALUNGA, MG | 137 |
| SFER, AM | 53 |
| SGUASSERO, Y | 68 |
| SICA, N | 124 |
| SILVA QUINTANA, S | 72, 140, 146 |
| SILVA, C | 142 |
| SIMONIELLO, MF | 72 |
| SMREKAR, M | 17 |
| SOFFIETTI, F | 97 |
| SOSA PAREDES, YK | 32 |
| SOSA-ESTANI, S | 68 |
| SOTO, J | 69, 70 |
| STANECKA, N | 120 |
| STEFANICH, CL | 154 |
| STIMOLO, MI | 43, 118 |
| STRUB, AM | 154 |
| SUAREZ, M | 142 |

| | |
|-----------------------|------------|
| T | |
| TABLADA, M | 67 |
| TALAY, C | 70 |
| TAPIA, E | 85 |
| TAUBER, L | 52, 60, 61 |
| TESSER, ME | 100 |
| TOLOSA, L | 111 |
| TONARELLI, G | 72 |
| TOPAYAN, M | 109 |
| TORRES MANNO, M | 89 |
| TORRES, CG | 136 |
| TORRES, PS | 35 |
| TRIANO, S | 57 |
| TRIPALDI, P | 115 |
| TROGLIERO, MC | 103 |
| TRUCÍOS, C | 50 |

| | |
|-----------------------------|-----|
| V | |
| VÁZQUEZ, MD | 19 |
| VADELL, D | 57 |
| VAIRA, SM | 102 |
| VALDIVIESO SERRANO, L | 32 |

VALENTINI, GH 92
VALENZUELA, C 128
VALLS, P 50
VARGAS, J 41, 43
VARGAS, JM 73
VASCONI, C 137
VECCHI, R 94
VELEZ, JP 76
VIANCO, AM 99
VICO, FD 148
VIDELA, ME 91
VIGNERA, ML 110
VILLAGRA, EL 147
VILLARREAL, F 86
VIOLLAZ, AJ 39

VITELLESCHI, MS 97, 104

W

WEIDMANN, G 39
WEITZ, D 156

Z

ZÚÑIGA, D 152
ZABALA, SM 106
ZACHARÍAS, D 75
ZANGIACOMI MARTINEZ, E 19
ZANUTIGH, V 83
ZEGA, B 124
ZERBATTO, M 102
ZILLONI, L 61
ZUBILLAGA, M 119