



XIV Congreso Latinoamericano de Sociedades
de Estadística “Laura Nalbarte”

Libro de Resúmenes

Organiza Sociedad Uruguaya de Estadística.



Sociedades que integran el Comité Científico



Apoyan



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Introducción

Durante los días 18 al 21 de octubre de 2021 se desarrolló de manera virtual con sede en Montevideo-Uruguay el XIV Congreso Latinoamericano de Sociedades de Estadística (CLATSE) denominado "Laura Nalbarte.^{en} homenaje a nuestra querida compañera quien falleció trágicamente el 14 de agosto de 2021, siendo, entonces, presidenta de la Sociedad Uruguaya de Estadística (SUE).

¿Qué es el CLATSE? ¹

A fines de la década de los 80 el directorio de la Sociedad Argentina de Estadística (SAE), fundada en 1952, y el de la Sociedad Chilena de Estadística (SOCHE) fundada en 1977, deciden potenciar sus congresos nacionales mediante la organización conjunta de congresos internacionales abarcando a las sociedades hermanas de América Latina dando lugar al CLATSE: eventos programados para celebrarse cada dos o tres años, alternando el país anfitrión. Es así que en 1991 se celebró el I CLATSE organizado por la Universidad Técnica Federico Santa María en Valparaíso, Chile. El II CLATSE fue organizado en 1993 por la Universidad Católica Argentina en Buenos Aires, Argentina. Siguiendo de manera itinerante entre estos dos países hasta que en 2008 la SUE organiza, junto con la Universidad de la República el VIII CLATSE en Montevideo, Uruguay. Continuando con el espíritu integrador y de difusión, en 2016 se incorpora Perú en la organización del XII CLATSE y en el 2018 México organiza el XIII CLATSE. En 2021 Uruguay vuelve a ser sede teniendo cómo desafío la organización del primer CLATSE virtual.

Organización

Cada año que se celebra el CLATSE, la Sociedad de Estadística del país anfitrión es la sociedad organizadora. Así, este año, el Comité Organizador estuvo a cargo de Laura Nalbarte (presidenta de la SUE hasta su fallecimiento el 14 de agosto de 2021), Ignacio Álvarez-Castro, Ana Coimbra, Leticia Debera, Miguel Galmés, Juan José Goyeneche y Eugenia Riaño, todos integrantes de la Comisión Directiva de la SUE. El Comité Científico estuvo conformado por integrantes de Sociedades de Estadística que conforman el CLATSE: por la AME (Asociación Mexicana de Estadística) Carlos Rodríguez y Silvia Rodríguez, por la SAE participaron Gabriela Boggio y Martín Saino; por la SOCHE Yolanda Gómez y finalmente por la SUE Natalia da Silva y Marco Scavino. Además de la integración del Comité Científico las Sociedades de Estadística del CLATSE participaron activamente en cuestiones organizativas a solicitud de la SUE.

Desarrollo

Más de 400 personas de 17 países del mundo participaron en la XIV edición del CLATSE, destacándose la participación de Argentina, Uruguay, México y Chile.

El Congreso contó con 5 actividades académicas: Cursos, Conferencias, Sesiones Invitadas, Sesiones de Ponencias y el "Premio Jorge Blanco".

Se dictaron 6 cursos: *Simulación de procesos demográficos en R* a cargo de Daniel Ciganda; *Geografía computacional con R* con Orlando Sabogal y Eugenia Riaño como docentes responsables; *Visualización de datos con R* a cargo de Natalia da Silva; *Introducción al análisis de sobrevivencia usando el software R* dictado por Diego Gallardo; *Aplicación de la estadística a la industria y Seis Sigma* a cargo de Humberto Gutiérrez Pulido y *Datos faltantes en control multivariado de procesos sobre componentes principales* que dictaron Marta Quaglino y Julia Fernandez.

Contamos con la presencia de 5 conferencistas de renombre. Abrió el congreso la PhD. Alicia Carriquiry cuya conferencia se tituló *Comparación de imágenes en aplicaciones forenses*; el Dr. Joachim Engel presentó *Enseñanza de las estadísticas para el empoderamiento y el compromiso social*, el Dr. Rogelio Salinas Gutiérrez presentó *Funciones de cópula y su aplicación en clasificación probabilística supervisada*; el Dr. Marcelo Ruiz *Selección de covarianza robusta para modelos gráficos Gaussianos en alta dimensión*; cerró el congreso el Dr. Mauricio Castro con su conferencia titulada *Análisis Bayesiano de datos de sobrevivencia con indicadores de censura faltantes*.

Con respecto a las Sesiones Invitadas, se formaron 6 mesas temáticas. Daniel Sánchez, Juana Sánchez y Hugo Hernández debatieron sobre *Los cambios de paradigmas de la presencialidad y la virtualidad en la*

¹Extraído de sue.org.uy

enseñanza universitaria de la Estadística, esta mesa estuvo a cargo de Adriana D'Amelio. El Director del INE de Uruguay, Ec Diego Aboal, el director del INDEC de Argentina, Dr. Marco Lavagna y el director de la División de Estadísticas de la FAO, José Rasero, disertaron sobre los *Desafíos de las Oficinas de Estadística en la era de Big Data*. *Avances recientes en la confiabilidad de sistemas complejos* fue el tema elegido por Diego Carvalho do Nascimento, Pedro Ramos, Paulo Henrique Ferreira da Silva. El Dr. Tomás Aluja, Dr. Javier Trejas, Mag. Federico Bayle intercambiaron ideas sobre *Ciencia de datos: una mirada desde la Estadística*. Graciela Muniz Terrera, Daniel Ciganda y Ramón Álvarez-Vaz disertaron sobre *Estadística aplicada a la salud y a las ciencias sociales*. Por último, integraron la mesa *Métodos estadísticos para datos de COVID-19*, el Dr. Martín Humberto Felix Medina, Dr. Mario Santana Ciprian, Dr. Netzahualcóyotl Castañeda Leyva.

Se presentaron 108 ponencias que fueron agrupadas en los siguientes temas:

- Tema 1: Estadística Bayesiana
- Tema 2: Bioestadística
- Tema 3: Estadística Computacional
- Tema 4: Datos faltantes
- Tema 5: Demografía
- Tema 6: Deporte
- Tema 7: Econometría
- Tema 8: Educación
- Tema 9: Enseñanza
- Tema 10: Espacial
- Tema 11: Análisis Longitudinal
- Tema 12: Machine Learning
- Tema 13: Métodos Estadísticos
- Tema 14: Misceláneos
- Tema 15: Muestreo
- Tema 16: Multivariado
- Tema 17: Salud

El presente documento incluye los resúmenes de las ponencias. Éstas fueron grabadas en un video de no más de 15 minutos cada una, videos que estuvieron disponibles días antes de comenzar el Congreso; las instancias de encuentro fueron espacios de intercambio por Zoom, donde se realizaron comentarios y preguntas a los autores.

Por último, con el apoyo económico del Instituto de Estadística de la Facultad de Ciencias Económicas y de Administración de la Universidad de la República, se entregó el premio a jóvenes investigadores, el Premio Jorge Blanco, al joven Dr Andrés Sosa por su trabajo sobre la modelización estocástica de las curvas de rendimiento de los bonos de deuda soberana de Uruguay.

Las grabaciones de las conferencias, las sesiones invitadas y la presentación del Premio Jorge Blanco están disponibles en el **canal de YouTube** del XIV CLATSE Laura Nalbarte.

Tema 1: Estadística Bayesiana

Is attention/focus related to sleep quality or a period of the day? A first look into the cognition of Learning.

Adolfo González González¹, Diego C. Nascimento²

¹*Departamento de Psicología, Universidad de Atacama (UDA), Chile.*

²*Departamento de Matemáticas, Universidad de Atacama (UDA), Chile.*

Nowadays lifestyle generates great demands (physically and mentally), whereas time seems to be not enough, causing the impression that days are shorter due the exhausting routines, varied or extensive tasks. Thus, features such as rest habits and nutritional supplementation are compromised with such routines. In this longitudinal pilot study, we analyzed 15 subjects, residents from Copiapó/Chile, during the month of June 2021, using a virtual questionnaire, containing 11 questions related to sociodemographic and lifestyle dimensions. Each participant was evaluated 5 times on the same day, for two consecutive days, respecting an interval of at least 3 hours between data collection adopting Stroop Color and Word Test (SCWT). Bayesian count regression was adopted to investigate the primes factors associated to the amount of right responses obtained from the SCWT. Results showed that a decrease of attention was notable around the lunch time (beginning of the afternoon), whereas some decrease of carbohydrate levels are to be expected, and affecting directly on the attention functionality.

Cuadro 1: Posterior summaries for Model 1

Parameter	Rhat	n_eff	mean	sd	2.5 %	25 %	75 %	97.5 %
(Intercept)	0.9994	1998	5.289	1.458	2.5113	4.3407	6.2069	8.2709
Sex	0.9994	1772	-0.065	0.259	-0.5511	-0.2290	0.0986	0.4836
Age	1.0007	2108	-0.023	0.012	-0.0456	-0.0304	-0.0160	-0.0002
Education	1.0015	1849	-0.024	0.065	-0.1614	-0.0651	0.0171	0.1025
Usual.sleep	1.0020	1740	0.04	0.134	-0.2182	-0.0391	0.1195	0.3102
Hours.needed	1.0009	1926	-0.063	0.124	-0.3181	-0.1415	0.0191	0.1831
Day	1.0007	2052	0.090	0.022	0.0483	0.0751	0.1043	0.1303
Hours.sleep	1.0004	2054	0.079	0.075	-0.0714	0.0297	0.1306	0.2240
Sleep.evaluation	1.0005	1833	0.006	0.017	-0.0270	-0.0051	0.0183	0.0377
Time.Period	0.9994	1941	0.013	0.014	-0.0143	0.0037	0.0227	0.0414
Lunch	0.9995	1905	-0.031	0.042	-0.1149	-0.0604	-0.0023	0.0513
Weighted.sleep.hours	1.0007	2092	-0.008	0.007	-0.0220	-0.0133	-0.0036	0.0067

Palabras Clave: Stroop Task, Sleep, Attention, Resistance to interference, Poisson regression.

¹adolfo.gonzalez@uda.cl

²diego.nascimento@uda.cl

Integración de Datos Heterogéneos de Gran Dimensión, Mediante Regresión Factorial Bayesiana

Alejandra Avalos-Pacheco¹, David Rossell, Richard S. Savage

Harvard-MIT center for regulatory science, Harvard Medical School, Harvard University

Dos desafíos clave en las aplicaciones estadísticas modernas son la gran cantidad de información registrada por individuo, y que esos datos a menudo no se recopilan todos a la vez, sino en lotes. Estos efectos por lotes pueden ser complejos y provocar distorsiones tanto en la media como en la varianza. Proponemos un nuevo modelo de regresión de factores latentes dispersos para integrar datos heterogéneos. El modelo proporciona una herramienta para la exploración de datos a través de la reducción de dimensionalidad y conjuntamente la corrección de efectos por lotes (batch effects). Estudiamos el uso de varias funciones de probabilidad a priori dispersas (locales y no locales) para inferir la dimensión de los factores latentes. Proporcionamos una metodología flexible para la regresión de factores espartos que no solo se limita a datos con efectos por lotes. Nuestro modelo se ajusta de manera determinista por medio de un algoritmo EM para el que derivamos actualizaciones de forma cerrada. Presentamos varios ejemplos, con enfoque en aplicaciones bioinformáticas. Nuestros resultados muestran un aumento en la precisión de la reducción de dimensionalidad, con distribuciones priori no locales (non-local priors) que mejoran sustancialmente la reconstrucción de la cardinalidad de los factores. Los resultados de nuestros análisis ilustran como no tener en cuenta adecuadamente los efectos por lotes pueden resultar en una inferencia poco confiable. Nuestro modelo proporciona un enfoque novedoso para la regresión de factores latentes que equilibra la escasez con la sensibilidad en escenarios con y sin efectos por lotes y es altamente eficiente computacionalmente.

Palabras Clave: Bayesian factor analysis, EM, non-local priors, shrinkage

¹e-mail: avalos@hms.harvard.edu

Rankings bayesianos: una aplicación de los estimadores de triple objetivo a datos de tiro de la Liga Uruguaya de Básquetbol

Ignacio Evangelista, Ignacio Álvarez-Castro¹

Facultad de Ciencias Exactas, Ingeniería y Agrimensura, Universidad Nacional de Rosario, Argentina
Instituto de Estadística, Universidad de la República, Uruguay

Las estadísticas de los deportes son una componente fundamental de la gestión de las instituciones deportivas. Con el paso de los años, la cantidad y la sofisticación de los modelos estadísticos aplicados a los deportes ha crecido y se ha potenciado. Los métodos bayesianos son cada vez más utilizados en el mundo de los deportes. Esta creciente popularidad se debe a que la inferencia bayesiana permite incorporar información experta *a priori*, aprendizaje secuencial, inferencias y predicciones probabilísticas, tratamiento de datos faltantes e inferencias con pocos datos, entre otras características.

Un objetivo natural del análisis de datos deportivos es el de ordenar o rankear los jugadores según habilidad. Cada jugador posee una habilidad latente para los tiros que se considera representada por un parámetro θ_k desconocido. Luego, dado el jugador k con habilidad θ_k , buscamos realizar inferencias sobre una cantidad no observable: su ranking dentro del conjunto Θ correspondiente a todos los jugadores. El ranking del jugador k , un valor $R(\theta_k) \in \mathbb{N}$, será el número de jugadores (más uno) que tienen que tienen una habilidad menor a este.

En el contexto bayesiano, los parámetros de habilidad se consideran variables aleatorias y por lo tanto tienen una distribución de probabilidad. Rankear los θ_k consiste en encontrar un ordenamiento a partir de las distribuciones de probabilidad *a posteriori*: $P(\theta_k | \mathbf{Y})$. El objetivo de estimar los rankings se contraponen con el de la estimación de los parámetros individuales θ_k : si se ordenan las medias *a posteriori* de las distribuciones de los θ_k , los rankings obtenidos no son óptimos.

Por otra parte, una característica que suele ser de interés es la función de distribución empírica o el histograma de los θ_k que permite, por ejemplo, calcular la fracción de jugadores que están por encima de un nivel. Tal como ocurría con los rankings, los mejores estimadores de los θ_k individuales inducen un histograma que no es óptimo.

Diferentes objetivos inferenciales dan origen a distintos estimadores. No obstante, puede ser relevante obtener un único conjunto de estimadores que se desempeñen bien para los tres objetivos: que produzcan un histograma que sea un buen estimador del histograma de los parámetros, que induzcan rankings que estimen apropiadamente el ordenamiento de los parámetros, y que además sean en sí mismos buenos estimadores de los parámetros. Shen y Louis (1998) propusieron los estimadores GR para cumplir con los tres objetivos. Estos estimadores han sido posteriormente utilizados en contextos de salud y educación.

En este trabajo se estudia la aplicación de los estimadores GR a un modelo jerárquico ajustado a los datos de tiro de la temporada regular 2020-21 de la Liga Uruguaya de Básquetbol. El número Y_k de tiros acertados por el jugador k en N_k intentos en la temporada se modeliza según una distribución binomial de parámetro θ_k . Se asume para los $\log\left(\frac{\theta_k}{1-\theta_k}\right)$ una distribución común normal de hiperparámetros μ y σ . Para μ y σ se elige como distribución *a priori* una forma conjugada débilmente informativa. Los estimadores GR se obtienen a partir de las muestras de MCMC obtenidas mediante **RStan** en R. Los resultados obtenidos se presentan mediante visualizaciones y se comparan los rankings con aquellos obtenidos mediante el ordenamiento de estimaciones puntuales (máxima verosimilitud y *medias a posteriori*).

Palabras Clave: inferencia bayesiana, estadísticas deportivas, rankings, modelos jerárquicos

¹e-mail: nachoeva@fceia.unr.edu.ar, nachalca@iesta.edu.uy

Inferencia Bayesiana para un modelo espacio-temporal usando la metodología INLA

Serrano Aguilar, Miguel Ángel¹

Departamento de Matemáticas, Universidad Autónoma Metropolitana - Iztapalapa

Con el desarrollo computacional el paradigma Bayesiano de la estadística ha cobrado gran importancia por su versatilidad para la inferencia de modelos complejos, sus buenos resultados y el desarrollo de técnicas y algoritmos computacionales eficientes como los MCMC (Gelman, 2006). Sin embargo, cuando se realiza inferencia con los métodos MCMC muchos recursos son utilizados al afinar y monitorizar la convergencia para encontrar el mejor ajuste; y cuando los modelos son complejos (especialmente cuando tienen un diseño jerárquico o cuando se utilizan bases de datos masivas) los métodos MCMC pueden ser extremadamente lentos o computacionalmente inviables.

Existen otras metodologías distintas a los MCMC con las que se puede hacer inferencia, una de ellas es el algoritmo de aproximación integrada anidada de Laplace o INLA (*Integrated Nested Laplace Approximation*), propuesto por Rue et al. (2009) que es un algoritmo determinista para inferencia Bayesiana. INLA está especialmente diseñado para modelos latentes Gaussianos y provee resultados precisos en un tiempo de cómputo menor a los MCMC (Gomez-Rubio, 2020).

Por otra parte, para ciertos modelos matemáticos se desea evaluar patrones espaciales así como temporales, este interés ha llevado al desarrollo de los modelos espacio-temporales. A estas variables se les puede asignar un componente aleatorio y describirse a partir de distribuciones de probabilidad, así el modelo descrito puede especificarse como:

$$Y(s, t) \equiv \{y(s, t); (s, t) \in \mathcal{D} \subset \mathbb{R}^2 \times \mathbb{R}\}$$

donde son observadas n localidades o áreas en T puntos de tiempo (Blangiardo, 2015). Estos modelos son usualmente definidos bajo estructuras jerárquicas Bayesianas.

En este trabajo ajustamos un modelo espacio-temporal utilizando, la metodología INLA, con un ejemplo de la literatura.

Palabras Clave: Inferencia Bayesiana, INLA, Modelos espacio-temporales, mapeo de enfermedades, incidencia de VIH

Bibliografía

- BLANGIARDO, M., CAMELETTI, M., *Spatial and Spatio-temporal Bayesian models with R-INLA*, Wiley, 2015.
- GAMERMANN, D., LOPES, H., *Markov Chain Monte Carlo. Stochastic Simulation for Bayesian Inference*, 2nd ed., Chapman & Hall/CRC Press, 2006.
- GÓMEZ-RUBIO, V., *Bayesian Inference with INLA*, CRC Press, 2020.
- RUE, H., MARTINO, S., CHOPIN, N., Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *Journal of the Royal Statistical Society, Serie B* 71(2), 319-392, 2009.

¹e-mail: m.serrano.aguilar@gmail.com

Una propuesta Bayesiana para medir el grado de traslape entre dos especies de animales

Gabriel Núñez Antonio¹

Departamento de Matemáticas, Universidad Autónoma Metropolitana, unidad Iztapalapa. México

Actualmente las técnicas de foto-trampeo, el desarrollo de modelos estadísticos más flexibles, el avance en la teoría sobre datos direccionales y la capacidad de procesamiento computacional de la información, han permitido un mayor desarrollo en ciertas áreas de las ciencias ecológicas. En particular, un mejor análisis de modelos para estudiar el traslape entre especies ofrece la oportunidad de entender de mejor manera los mecanismos de coexistencia de ciertos animales en su habitat natural.

Con el objetivo de contribuir al estudio del traslape entre especies de mamíferos a través de datos de foto-trampeo, en este trabajo se analiza un modelo bayesiano no paramétrico para variables circulares. Se propone una estimación de las densidades predictivas para los registros de avistamientos de animales y se ofrecen inferencias sobre una medida de traslape propuesta originalmente por Weitzman (1970).

Palabras Clave: Estimación de densidades, Proceso Dirichlet, Inicial de referencia no paramétrica.

¹e-mail: gab.nuneza@gmail.com

Tema 2: Bioestadística

Comparación de diferentes modelos de GWAS para una base de datos de porotos (*Phaseolus vulgaris* L.).

BORTOLOTTO E. B.¹ ANGELINI J., FAVIERE G. S., BRUNO C., CERVIGNI G.

Centro de Estudios Fotosintéticos y Bioquímicos (CEFOBI), Universidad Nacional de Rosario - CONICET, Suipacha 531, 2000 Rosario, Argentina.

El objetivo de este trabajo fue evaluar modelos estadísticos en cuanto a su desempeño para estimar asociaciones del todo el genoma (GWAS) entre el fenotipo y genotipo. El fenotipo fue medido sobre la floración medida como día hasta la primera floración (DFP); días de floración (DTF) y días hasta la última floración (DEF). Se trabajó con un conjunto de datos compuesto por 80 cultivares de porotos genotipados con 380 marcadores moleculares del tipo SNP (*Single Nucleotide Polymorphism*). Los valores fenotípicos fueron previamente ajustados por un modelo lineal mixto considerando el efecto del ambiente como fijo y al efecto del cultivar e interacción de ambos como aleatorios siguiendo Nascimiento et al. (2018). Previo al análisis GWAS se realizó un análisis de diversidad de los SNP, a través de métodos multivariados como Análisis de Componentes Principales y Análisis de Conglomerado (UPGMA y Bayesiano) y se buscó la matriz de estructura poblacional Q. Según el análisis UPGMA se pudo armar 7 grupos de marcadores, y según ACP esta cantidad de componentes explican el 82% de la variabilidad total de los marcadores. Los modelos comparados fueron: ANOVA: Modelo Lineal General, Modelos Lineal General con componentes principales (GLM-P), Modelos Lineal Mixto con PCA y matriz Kinship (MLM-PK), MLM comprimido (CMLM), CMLM enriquecido (ECMLM), unificación de probabilidad circulante de modelo fijo y aleatorio (FarmCPU) y Regresión por Cuantiles (QR).

Los modelos comparados fueron evaluados a través de gráficos de cuantiles (Q-Q) del logaritmo de los valores p observados en función de los esperados. Si los valores p obtenidos del modelo ajustado se ubican por encima de la línea recta a 45° indica la presencia de falsos positivos, es decir marcadores moleculares cuyo valor indica que existe una asociación estadísticamente significativa entre el genotipo y el fenotipo, pero que bajo hipótesis nula no lo son. Por el contrario si los valores p de los modelos se ubican por debajo de la línea recta, indican falsos negativos, es decir, marcadores SNP que están asociados al fenotipo pero su valor p no fue estadísticamente significativo. Si los valores p obtenidos se encuentran sobre la recta entonces indican que existen asociaciones verdaderas y polimorfismos causales. Para la corrección por multiplicidad se utilizaron dos métodos: Bonferroni y tasa de falsos descubrimientos (FDR).

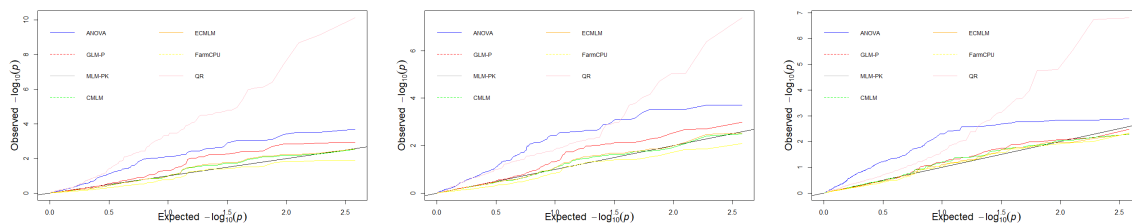


Figura 1. Gráfico de cuantiles (Q-Q) de los 7 modelos GWAS ajustados para 80 cultivares de porotos genotipados con 380 marcadores tipo SNPs, fenotipados para Días desde la primera Floración (DFP-izquierda), Días desde floración (DTF-centro) y días hasta última floración (DEF-derecha).

Basados en los gráficos Q-Q plot los modelos FarmCPU y CMLM fueron los que mejor comportamiento tuvieron respecto a los otros dado que sus curvas se ubican cercanos al a recta 1:1. En este sentido, el modelo que mayor tasa de falsos positivos presentó fue QR.

Palabras Clave: Modelos de Asociación, fenotipo, genotipo, marcadores SNP

¹e-mail: bortolotto@cefobi-conicet.gov.ar

Manejo bioestadístico de recursos fitogenéticos en la caracterización morfológica de accesiones de cañihua (*Chenopodium pallidicaule* Aellen), refrescadas en la EEA. Illpa, Puno Perú.

José Luis Cahuana Jorge¹

Centro de Investigaciones Avanzadas y Formación Superior en Educación, Salud y Medio ambiente 'AMTAWI'. Puno, Perú.

Conservar los recursos genéticos vegetales a nivel de banco de germoplasma, es uno de los actuales retos de la ciencia. Sin embargo, a pesar de múltiples esfuerzos realizados en diversas investigaciones, los resultados en más de una ocasión, no son visibles para su introducción y generalización social, dado que aún persisten limitaciones en el manejo de la bioestadística, lo cual no permite tomar decisiones prácticas para la conservación de recursos filogenéticos. La cañihua (*Chenopodium pallidicaule* Aellen), es uno de cultivos andinos que carece de registro analítico formal, dada su diversidad morfológica principalmente en la Región Puno; para ello, se realizó este estudio en el banco de germoplasma de la Estación Experimental Agraria de Illpa (perteneciente al Instituto Nacional de Innovación Agraria: INIA) en la propia provincia de Puno, la caracterización se hizo en dos grupos de variables, tanto cuantitativas como cualitativas relacionadas con la morfología y fisiología del cultivo. Para el análisis de los descriptores de las variables de interés, fueron aplicados métodos paramétricos con estadígrafos, así como análisis multivariados, donde finalmente se pudo observar la característica de la diversidad en el germoplasma.

En cuanto a los resultados se evidencia que las accesiones de cañihua presentan comportamientos variados en su desarrollo; tal es así, que se encontró un 47% de las accesiones presentaron poca germinación y posteriormente un 66% en fase fenológica de grano pastoso, en cuanto al color de tallo, inicialmente el 47% eran de color verde y que en la segunda evaluación presentaron una coloración de rojo claro. En cuanto al análisis multivariado, las dos primeras componentes explican el 77% de variabilidad de las accesiones de cañihua, donde la mayor contribución en el primer componente está dado por la variable “altura de planta” y en la segunda por “diámetro de tallo2”. En cuanto al análisis cluster se ha logrado obtener 3 grupos diferenciados, las mismas que han sido refrendados por un análisis de varianza que permite diferenciarlos con un $p - valor < 0,05$.

Palabras Clave: Bioestadística, cañihua, caracterización, germoplasma, recursos fitogenéticos.

¹e-mail: jose1cj@gmail.com

Aplicación de modelos de conteo para los registros del agutí (*Dasyprocta* sp.) en cercanías al Parque Nacional Calilegua (Jujuy, Argentina)

Valladares G.A.(1), Albanesi S.A.(2)(3), Cáceres F.(2), Arce O.E.A.(1)¹

(1)*Cátedra de Biometría y Técnica Experimental, Facultad de Agronomía y Zootecnia, Universidad Nacional de Tucumán, Argentina*

(2)*Fundación Pro Yungas*

(3)*Instituto de Biodiversidad Neotropical, CONICET-Universidad Nacional de Tucumán, Argentina*

El agutí (*Dasyprocta* sp.) es un pequeño roedor que habita las Yungas de Argentina, encontrándose en las provincias de Salta y Jujuy. Es una especie considerada ecológicamente importante para la dinámica del bosque porque dispersa las semillas de los frutos consumidos. Es una especie abundante que no se encuentra amenazada y habita principalmente lugares en buen estado de conservación encontrándose también en sitios degradados. Dentro de las campañas de monitoreo de mamíferos con la metodología de cámaras trampa que se llevan a cabo desde el año 2009 en cercanías al Parque Nacional Calilegua (Jujuy), se identificaron tres sitios de muestreo relevantes. Se definieron tales sitios de acuerdo con la accesibilidad (capacidad del hombre en llegar al lugar ya sea en busca de alimento, madera o con fines cinegéticos) y se establecieron así los sitios de baja, media y alta accesibilidad. Se colocaron diez cámaras trampa por cada sitio de muestreo en los años 2009, 2014 y 2019.

El objetivo de este trabajo fue modelar las visualizaciones de ejemplares de agutí en 3 años de observaciones y 3 tipos de accesibilidad mediante modelos para datos de conteos. Para el análisis de los datos consideramos como registros independientes a aquéllos que ocurrieron en intervalos de más de 12 horas entre fotos sucesivas de la misma especie. La variable respuesta fue el conteo de observaciones de agutí. El esfuerzo de muestreo, que fue diferente en los distintos puntos de registro, se utilizó como covariable. Para el análisis de los datos se usaron modelos para datos de conteo: modelos lineales generalizados de las familias Poisson y Binomial negativa y modelos inflados en ceros de las mismas familias. El ajuste se evaluó mediante el criterio de información de Akaike (AIC) y análisis de residuos. Los modelos de Poisson no lograron modelar la sobredispersión de los datos, proporcionando los valores de AIC más altos. Ambos modelos de la familia binomial negativa manejaron la sobredispersión generando valores de AIC casi idénticos. Se decidió por el modelo sin inflación en ceros ya que sus residuos tuvieron mejor comportamiento. La interacción año x accesibilidad del sitio resultó no significativa por lo que fue excluida del modelo. Para el efecto año, se observó un 46 % menos de observaciones en 2014 con respecto a 2009, no habiéndose encontrado diferencias significativas entre 2009 y 2019. Respecto a la accesibilidad de los sitios, el conteo fue 1,66 veces mayor en el sitio de accesibilidad media que en el de accesibilidad baja. Esto puede deberse a que el agutí es una especie que generalmente está asociada a cursos de agua y el sitio de accesibilidad media es un sitio más húmedo en comparación con los demás ya que posee varios cursos de agua importantes y permanentes. Con respecto a los sitios con alta y baja accesibilidad, el conteo del agutí fue un 48 % menor en los sitios con alta presencia humana. Esto puede deberse a que poseen una historia de altas frecuencias de incendios, de extracción de madera ilegal y a que pueden funcionar como entrada para cazadores que ante el encuentro con cualquier animal silvestre también pueden cazarlo, aún sin tener el agutí un alto valor cinegético.

Palabras Clave: Bioestadística, Modelos de conteo, Agutí, Modelos lineales generalizados

¹e-mail:osvaldo.arce@faz.unt.edu.ar, valladares.gabriela@gmail.com, sebastianalbanesi@gmail.com, robertocaceres@proyungas.org.ar

Decodificando información en redes multipartitas: Usando clusterización y teoría de la información

Huaylla Claudia^a, Kuperman Marcelo^b, Garibaldi Lucas^{a 1}

- a. Consejo Nacional de Investigaciones Científicas y Técnicas. Instituto de Investigaciones en Recursos Naturales, Agroecología y Desarrollo Rural. Universidad Nacional de Río Negro. San Carlos de Bariloche, Río Negro, Argentina.
- b. Centro Atómico Bariloche (CNEA) and CONICET. Instituto Balseiro, Universidad Nacional de Cuyo, R8402AGP Bariloche, Argentina.

El análisis de un ecosistema puede ser abordado usando diferentes herramientas, una de ellas son las redes complejas. Las redes permiten capturar la estructura de la topología definida por las interacciones entre las especies. El análisis de sus propiedades constituye una forma clarificadora de entender el funcionamiento del ecosistema. Muchos trabajos no se detienen a estudiar si la red presenta información y realizan directamente el análisis de la misma. Proponemos un método para determinar si una red tiene información relevante en su sistema. Este método consiste en aleatorizar la red original intercambiando enlaces y conservando el grado de cada nodo. Si la red presenta información en su sistema, una aleatorización secuencial mostraría que los valores del índice de modularidad decrecen cuando incorporamos cambios en la red. Este método fue aplicado a una red trófica de tres niveles, polinizador, planta y herbívoro. Nosotros encontramos que la modularidad decrece al aumentar el desorden. Se realizó un test de hipótesis para determinar si el valor de la modularidad de la red original era significativamente distinta del valor obtenido después de realizar 80 cambios, es decir, cuando un valor estable de la modularidad fue alcanzado y se probó que a nivel 0.05 los valores eran significativamente distintos.

Además, calculamos la información mutua y la variación de la información en las redes derivadas de la red original usando 3 algoritmos de clusterización (Louvain, Newman-Girvan y Spectral Clustering). La información mutua cuantifica la pérdida con respecto a la clusterización original (C) y la variación de la información es la suma entre la cantidad de información que perdemos sobre la clusterización C y la cantidad de información que ganamos cuando pasamos de C a C' . La información mutua comenzó a decrecer a medida que se incrementaron los cambios y en la variación de la información se observó una tendencia creciente.

Nuestros resultados muestran que el análisis basado en la medida de modularidad es una herramienta confiable para caracterizar redes y que la información mutua y la variación de la información son herramientas para cuantificar la pérdida de información.

Palabras Clave: Bioestadística, ecosistema, información mutua, modularidad, redes aleatorias

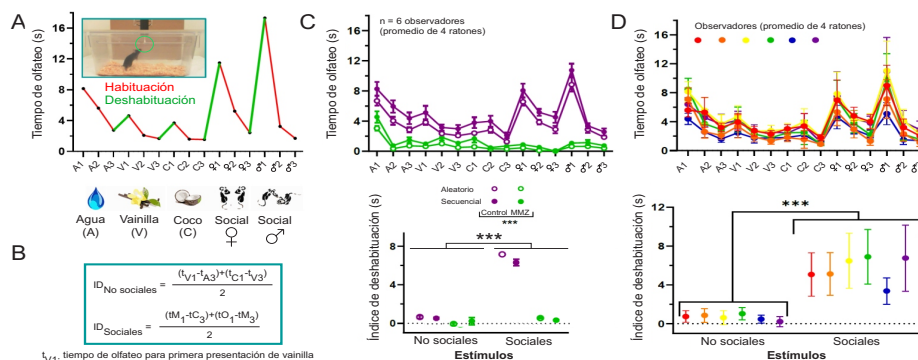
¹e-mail: cahuaylla@unrn.edu.ar, kuperman@cab.cnea.gov.ar, lgaribaldi@unrn.edu.ar

Evaluating possible sources of bias when administering the olfactory habituation-dishabituation test in mice

Diego M Topsakalian, Javier H Fotti, Lorena Rela y Gilda Garibotti¹

*IFIBIO Houssay, Facultad de Medicina, Universidad de Buenos Aires
Departamento de Estadística, Universidad Nacional del Comahue Bariloche*

En modelos de daño reversible al nervio olfatorio, la integridad funcional refleja la recuperación del daño y puede evaluarse con la prueba de habituación-deshabituaación a olores. La exposición repetida a un estímulo reduce progresivamente la actividad de olfateo (habituaación), mientras que se incrementa el tiempo de olfateo ante un odorante novedoso (deshabituaación). La conducta se parametriza utilizando índices de deshabituaación (ID, Figuras 1A-B). El tiempo de olfateo se cuantifica manualmente a partir de videos. Los observadores son ciegos a los tratamientos, pero la evaluación es susceptible de sesgos, como aquellos por expectativa y por deriva. El sesgo por expectativa se da si una hipótesis influye la cuantificación. Para evaluarlo, reclutamos 12 voluntarios, que cuantificaron la conducta olfativa de ratones normósquicos y anósquicos. Los videos de la conducta se fragmentaron en cada presentación de estímulo y recompilaron secuencial o aleatorizadamente (con presentaciones de diferentes estímulos, tratamientos y/o animales). Mediante ANOVA (modelos mixtos), no detectamos efecto significativo de la aleatorización ($p = 0,9095$), ni interacción significativa con tratamiento ($p = 0,3057$), con estímulo ($p = 0,5969$), o con ambos ($p = 0,0783$) (Figura 1C). Así, en nuestras condiciones, aleatorizar no ofrece ventaja. La deriva del observador se refiere al cambio en los criterios de análisis con el tiempo. Para su evaluación, comparamos datos obtenidos por observadores que analizaron videos de ratones normósquicos, compilados secuencialmente. No observamos diferencias significativas entre dos modelos que diferían en la inclusión del observador como variable aleatoria ($p = 0,9999$). Inferimos que entre el entrenamiento de los observadores y la finalización del análisis de la conducta no se produjo deriva que afecte la reproducibilidad interindividual (Figura 1D). Finalmente, para reducir la variabilidad, evaluamos el efecto de asistir visualmente al observador mediante una región de interés (RdI) dentro de la que el olfateo debe ser cuantificado (Figura 1A). Se comparó la cuantificación de la conducta de ratones normósquicos en una serie de presentaciones ($n = 8$) observadas con la RdI delimitada o sin ella, cada una por duplicado y en orden aleatorio. La RdI no afectó significativamente al tiempo de exploración promedio ($p = 0,6552$) ni se observó interacción entre presencia de RdI y estímulo ($p = 0,7607$). En cambio, se vio un efecto significativo de la presencia de RdI en la reproducibilidad intraindividual ($p = 0,0193$). Los resultados permiten recomendar que la cuantificación se haga por observación secuencial de la conducta y con RdI guía. Además, las cuantificaciones de diferentes observadores pueden interpretarse conjuntamente.



Palabras Clave: olfato, ratones, conducta, sesgos, reproducibilidad, habituación, deshabituaación

¹e-mail: diegotopsa@gmail.com, javier.fotti@gmail.com, lorena.rela@fmed.uba.ar, garibottig@comahue-conicet.gob.ar

Alternativas metaanalíticas para la determinación de modelos genéticos

Cuesta Cristina 1, Petroni Candela 2, Martinelli Romina 3-4, Sguassero Yanina 5¹

1 Instituto de Investigaciones Teóricas y Aplicadas en Estadística, Fac. de Cs. Econ. y Estad., UNR, 2 Cátedra de Fisiología Humana, Fac. de Cs. Médicas, UNR, 3 CONICET, 4 Cátedra de Genética Humana, Fac. de Medicina y Cs. de la Salud, UAI, 5 Investigadora independiente, revisora sistemática

Las revisiones sistemáticas (RS) son documentos de síntesis de la evidencia científica en medicina que permiten la búsqueda, identificación y apreciación crítica de todos los estudios primarios que dan respuesta a misma pregunta. En ese contexto, los metaanálisis constituyen el cuerpo de técnicas estadísticas que se llevan a cabo para determinar una estimación global a partir de los estudios individuales incluidos en una RS.

En los últimos años, las RS sobre asociaciones entre polimorfismos genéticos y ciertas enfermedades han ido ganando relevancia y también lo han hecho las técnicas de metaanálisis asociadas al tema.

En un metaanálisis tradicional, se comparan estudios que tienen dos grupos (por ejemplo tratado/control) y a partir de ellos se obtiene un efecto global de asociación entre los mismos y el evento relacionado con la enfermedad. En los estudios de asociación genética hay al menos tres grupos (si los alelos son A y a, los genotipos serán AA, Aa, aa).

En este escenario, se han desarrollado diferentes alternativas metaanalíticas para establecer el modelo genético subyacente (dominante, recesivo, dominante completo o codominante) asociado con el evento de interés. Previamente al metaanálisis, es necesario probar que se cumpla el equilibrio de Hardy-Weinberg en cada estudio, a través de un test de bondad de ajuste.

Thakkestian propone que, luego de analizar la homogeneidad de los genotipos de a pares, se pruebe la significación de los genotipos a partir de un modelo de regresión logística controlando por el efecto del estudio. En caso de que el efecto genotipo resulte significativo, se analizan las razones de odds de AA vs aa, Aa vs aa y AA vs Aa. A partir de ellas se concluye sobre el modelo genotípico, se colapsan los genotipos, se vuelve a chequear la homogeneidad y en caso de satisfacerse, se estima el efecto de asociación global.

Por su parte Minelli, presenta otra alternativa de análisis para polimorfismos bialélicos basado en un modelo de regresión lineal donde se relaciona el logaritmo de la razón de odds de AA vs aa con el logaritmo de la razón de odds de Aa vs aa. La pendiente de ese modelo es un indicador del modelo global.

El objetivo de este trabajo es comparar estas dos alternativas metaanalíticas en el marco de una RS de estudios de casos y controles que ha sido realizada para explorar la asociación entre los polimorfismos del receptor de vitamina D (BsmI, ApaI, TaqI y FokI) y el cáncer de próstata.

Palabras Clave: bioestadística, metaanálisis, modelos asociación genética, revisiones sistemáticas

¹e-mail: ccuesta@fcecon.unr.edu.ar

Modelo Binomial Negativo Mixto para el análisis de datos con sobre dispersión en susceptibilidad a la infección por *Stemphylium vesicarium* en frutales del Alto Valle

BASANTA M.,¹ GIMENEZ G.1, RUBIO, N.1, TUDELA A.2 ¹

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

El presente estudio tuvo como objetivo determinar la susceptibilidad de los principales cultivares de pera presentes en la zona productiva del Alto Valle a la infección por *Stemphylium vesicarium*. Conocer la susceptibilidad de los frutos de las distintas variedades a lo largo de su desarrollo resulta útil a la hora de considerar las medidas de control de la enfermedad. Los frutos cosechados se desinfectaron convenientemente y luego de serles inoculados con una cepa altamente patogénica, identificada como S23, se dispusieron dentro de cámaras húmedas construidas por una bandeja con un soporte individual para cada fruto. Para la determinación de la incidencia de la enfermedad, a los 7 días de incubación, se contabilizó el número de manchas por fruto. Dada la naturaleza de la variable respuesta “número de manchas” se ajustó un modelo Poisson en función de la variedad del fruto y el momento de cosecha del mismo. Debido a que los datos presentaban sobredispersión se utilizó un Modelo Binomial Negativo el cual mejoró notablemente el ajuste del modelo. Pero debido a que el mismo fruto se midió a lo largo del tiempo es necesario tener en cuenta el efecto del fruto como individuo dentro del modelo y para ellos se planteó un Modelo Binomial Negativo Mixto. De acuerdo a los resultados obtenidos, se puede afirmar que la edad de fruto tiene influencia sobre la susceptibilidad a la “Mancha marrón del peral”, y, que el efecto de la variedad es muy importante, ya que es determinante en el desarrollo de lesiones y el avance de las mismas.

Modelo Lineal Generalizado Mixto, *Stemphylium Vesicarium*, Sobredispersión, Datos de conteo, número de lesiones:

¹e-mail: mabel.basanta@faea.uncoma.edu.ar

Uso de las Regresiones Lasso en un Modelo Binomial Negativo para analizar la presencia del Chimango Patagónico en el Aeropuerto de la Ciudad de Neuquén

RUBIO N.1, GIMENEZ G.1, BERNARDIS A. 1, ANGELICCHIO J 1¹

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

El objetivo de este trabajo es determinar qué factores climáticos afectan a la presencia de aves en el Aeropuerto Internacional de la ciudad de Neuquén. En particular se estudió la presencia del chimango Patagónico dado que el mismo significa un peligro avial para el tráfico aéreo. Diversos estudios han estudiado el comportamiento de estas aves pero ninguno se ha encargado de determinar qué factores climáticos favorecen la presencia de Chimango por lo tanto en este estudio nos enfocamos en determinar cuáles son los factores/ variables que explican la presencia de estas aves para así lograr un modelo predictivo. Para ello se registró la frecuencia de chimangos semanalmente mediante el método de conteo por puntos durante 1 hora al amanecer y las principales variables climáticas como velocidad del viento, nubosidad, época del año, temperatura máxima y mínima del día, humedad y precipitación promedio. Considerando la variable de respuesta cantidad de Chimangos avistados por día, se aplicó en primera instancia un modelo Poisson pero al determinar la falta de ajuste debido a la sobredispersión de los datos se aplicó un Modelo Binomial Negativo. Comenzando el análisis con un modelo saturado para mediante el uso del algoritmo stepwise arribar a un modelo más parsimonioso que tiene en cuenta solo a los efectos principales tales como: la humedad, la velocidad del viento y Estación. Al mismo tiempo se aplicó una regresión Lasso para la selección de variables. Luego se comparó la Deviance Residual de los modelos obtenidos con ambas técnicas para el modelo Binomial Negativo. Debido a los problemas de convergencia del modelo ya que la cantidad de individuos es inferior a la de variables es que se plantea el uso de las Regresiones Lasso para la variable cuya distribución es Binomial Negativa. Una de las ventajas para la selección de variables mediante el uso de las Lasso es que se puede aplicar en un contexto de altadimensionalidad mientras que en MLG no se logran estimar todos los efectos cuando se cuentan con los suficientes individuos. Al finalizar el proceso de selección tanto para la variable considerada como Poisson como la Binomial Negativa se obtuvo el modelo que consideraba las interacciones dobles entre la Humedad, la Velocidad del Viento y la Estación del Año. Luego se analizó la Deviance Residual y los valores del AIC de éste modelo con el que contenía sólo dichos efectos principales observando que el modelo que consideraba a la variable de respuesta Binomial Negativa y que sólo contaba con los efectos principales era el adecuado (Modelo 1: $C + V + H + E$). A continuación se utilizó el proceso de selección mediante el uso de Regresión Lasso considerando la distribución Binomial Negativa para la cantidad de chimangos avistados por día arribando al siguiente resultado: Modelo 2: $C + E + t_{min} + h_{min} + V + NP + H^2 + t_{max}^2 + t_{min}^2 + D^2$. Finalmente se compararon ambos modelos (modelo 1 y 2) mediante el análisis de la deviance residual como los valores del Pseudo R².

Por lo que se puede arribar a las siguientes conclusiones: Una de las ventajas en el uso de la Regresión Lasso es que se puede aplicar en un contexto de altadimensionalidad ($n \ll p$). Una limitación computacional de las Lasso es que no considera interacciones ni modelos polinómicos, lo cual se puede resolver mediante la construcción de la matriz de diseño. Ambos modelos explicaron la cantidad de chimangos con las variables Estación, Humedad y Velocidad del Viento. Además fueron seleccionadas con Lasso Día², Temperatura Mínima² y Máxima² y Humedad Máxima². Se compararon el PseudoR² y la Deviance de los modelos obtenidos con ambas técnicas. Resultando menor para el modelo seleccionado con Stepwise y ambos Pseudos R² son altos lo cual indica que ambos modelos son buenos ajustes.

»Regresión Lasso, Modelo Lineal Generalizado, Conteo por Puntos, Chimango Patagónico, Selección de Modelos

¹e-mail: natalia.rubio@faeauncoma.com.ar

ANÁLISIS DE LA VARIABILIDAD CONJUNTA DE REMOCIÓN DE CONTAMINANTES EN AGUA RESIDUAL URBANA POR HUMEDALES CONSTRUIDOS

Stella Vaira, Mariel Zerbato, Ana Pizarro, Laura Modini (Directora del Proyecto)¹

Facultad de Bioquímica y Ciencias Biológicas: LISEB; Cátedra de Tratamiento de Efluentes.
Universidad Nacional del Litoral

Los humedales artificiales (HA) son sistemas diseñados y construidos para aprovechar los procesos naturales que involucran macrófitas, medio filtrante y microorganismos en el tratamiento de aguas residuales (Vymazal, 2014). Debido al bajo costo de instalación, operación y mantenimiento, representan una alternativa atractiva para el tratamiento de efluente urbano en pequeñas comunidades. Este estudio investiga la eficiencia de HAs de flujo subsuperficial vertical (HFVSS) para depurar agua residual urbana. Se construyeron 3 humedales con *Schoenoplectus californicus* y distintos lechos: Ha (arena gruesa), Hc (coque), Ec (arena y coque). La alimentación se realizó en forma continua con efluente cloacal real. El tiempo de residencia hidráulica fue de 48 h. La eficiencia de remoción de contaminantes depende de diversos factores, como el diseño, los materiales del lecho y el régimen de alimentación. En este trabajo se utilizan técnicas multivariadas para analizar el patrón de variación de diferentes variables contaminantes, durante el período estudiado (11 semanas). En particular se utilizó PCA (Principal Component Analysis) algoritmo de aprendizaje automático no supervisado que intenta reducir la dimensionalidad (número de características) dentro de un conjunto de datos mientras retiene toda la información posible. Variables que se incluyeron: X1: % de Remoción de Demanda Química de Oxígeno (mg/L), X2: % de Remoción de Turbiedad (UNT), X3: % de Remoción de Conductividad ($\mu\text{S}/\text{cm}$), X4: % de Remoción de Amoníaco (NH_4^+/L), X5: % de Remoción de Color, X6: % Remoción FRT (Fósforo Reactivo Total), X7: % Remoción PRS (Fósforo Reactivo Soluble), X8: % Remoción Sólidos Suspendedos Totales, X9: Formación de Nitrato ($\text{mg NO}_3^-/\text{L}$). Se compararon la eficiencia de los parámetros en los 3 HA mediante el test de Kruskal – Wallis. La distancia de Mahalanobis se calculó como medida para la detección de outliers multivariado. Se utilizó la librería FactoMineR y los paquetes biplot(), princomp(), kruskal.test() (R Core Team, 2018). Nivel de significancia α prefijado fue de 0,05. El número óptimo de componentes resultó en cuatro. La Figura muestra la ordenación (PCA) de las variables contaminantes involucradas en el análisis de los $n = 25$ datos.

Las dos primeras componentes acumulan un 53,1% de la varianza explicada, las primeras 4 componentes un 81,3%. Un análisis cualitativo conduce a separar los parámetros en remoción de nutrientes y sólidos disueltos de la materia orgánica y sólidos suspendidos.

Las comparaciones de las variables contaminantes para los diferentes humedales construidos, resultaron con diferencias significativas: Remoción de Turbiedad y Formación de Nitrato ($p < 10^{-3}$). El mayor porcentaje de remoción de Turbiedad se encontró en el sistema Ec, en ese mismo sistema el mayor valor de Nitrato. Para las variables X6 y X7, los valores p resultaron 0,053 y 0,037 respectivamente, alcanzando sus máximos valores de remoción en el sistema Ha. Los tres sistemas eliminaron % DQO con eficiencias similares $78,3 \pm 9,4$, como media y desvío estándar global. La distancia de Mahalanobis resultó de 4,91 y no se presentaron valores atípicos.

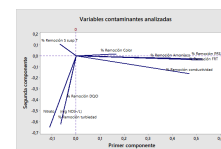


Figura. Componentes 1 y 2 para las variables contaminantes, 3 HA.

Vymazal, J. (2014). Constructed wetlands for treatment of industrial wastewaters: A review. *Ecol. Eng.*, 73 : 724 – 751.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Palabras Clave: humedales, contaminantes, remociones, análisis multivariado.

¹e-mail: stella.vaira@mail.com, mzerbato@fbc.unl.edu.ar

Tema 3: Estadística Computacional

Elección de la red neuronal adecuada para la estimación de la demanda de carga en sistemas de distribución de fluido eléctrico.

Marcelo Smrekar, Agustín Martina, Gonzalo Moya, Andrés Rodríguez, Luis Giovine¹

Laboratorio de Hidráulica, Facultad de Ciencias Exáctas Físicas y Naturales, Universidad Nacional de Córdoba.

El estudio de la demanda eléctrica y de su predicción tienen gran cantidad de antecedentes en la bibliografía tanto en las estrategias de estimación como en las técnicas y modelos para la predicción. La carga del sistema eléctrico es un proceso aleatorio continuo, con muchas variables e influenciado por varios factores. Parte de esta investigación es justificar el método de pronóstico elegido para realizar la estimación de la serie de demandas. Éstas presentan diferentes niveles de estacionalidad y cuentan con numerosas variables exógenas involucradas: clima, cambios tecnológicos y sociales que producen quiebres y cambios de patrones en las series. La estrategia de modelado fue evolucionando desde los modelos paramétricos lineales pasando por los paramétricos no lineales hasta llegar a los modelos no paramétricos de la estadística computacional. Últimamente se han popularizado los modelos de aprendizaje artificial, aprendizaje profundo e inteligencia artificial. En líneas generales, hay consenso en la bibliografía en que el método de Redes Neuronales Artificiales (RNA) es adecuado para la estimación de la demanda de carga.

En este trabajo se justifica la elección de la herramienta estadística para predecir la demanda y se realiza una optimización para comparar el desempeño de distintas topologías de RNA, distintas arquitecturas, distintas fuentes de datos meteorológicos y la incorporación de distintas variables meteorológicas en la predicción de la demanda eléctrica con RNA utilizando los paquetes Keras y Tensorflow de Python.

Palabras Clave: Redes Neuronales- Aprendizaje Artificial- Estimación- Series Temporales.

¹e-mail: marcelo.smrekar@unc.edu.ar, amartina@unc.edu.ar, gonzalo.moya@gmail.com, androadminplan@gmail.com, ingluisgiovine@gmail.com

Librería *raschreg* para modelos de teoría del respuesta al ítem en contextos de regresión

Fernando Massa, Ramón Álvarez-Vaz¹

Instituto de Estadística, Departamento de Métodos Cuantitativos, Universidad de la República

Se presenta una versión preliminar de la librería *raschreg* de R, la cual está pensada para extender los modelos de teoría del respuesta al ítem (TRI) (de uno y dos parámetros) de manera de incluir variables explicativas del parámetro individual. La estimación de los modelos se lleva a cabo dentro de un marco de máxima verosimilitud marginal aproximando de las integrales mediante la aproximación de Laplace.

La especificación de la distribución del parámetro individual es normal estándar y puede incluir o no variables explicativas en la media de dicha distribución. Se presentan las primeras funcionalidades de la librería, a través de diferentes ayudas (vignettes) que, junto con la librería, se encuentran disponibles en el repositorio de Github <https://github.com/nando11235813/raschreg>. Se presenta la estimación de los modelos logísticos de 1 y 2 parámetros, mostrando en ambos casos la posibilidad de incluir variables de regresión. Se consignan visualizaciones de los resultados de dichos modelos a través de curvas características del ítem, la función e información, los mapas de personas-ítem y forest-plots.

Finalmente se presenta una aplicación del área de la epidemiología usando estos modelos para determinar la salud de un conjunto de personas que demandaron atención en el Servicio de registros de la Facultad de Odontología de Udelar, empleando como ítems la presencia (o ausencia) de diferentes patologías y como variables explicativas variables sociodemográficas y factores comportamentales.

Palabras Clave: Modelos Mixtos Generalizados, Salud bucal, Teoría de respuesta al ítem, R

¹e-mail: fmassa@iesta.edu.uy, ramon.alvarez@fcea.edu.uy

Estimación de un modelo computacional mediante ABC

Juan Ignacio Baccino, Mauro Loprete¹

*Estudiantes de la licenciatura en Estadística,
Facultad de Ciencias Económicas y de Administración - Udelar,
Montevideo, Uruguay*

Daniel Ciganda²

Max Planck Institute for Demographic Research

Resumen

El primer objetivo de este trabajo es estimar los parámetros de un modelo computacional del comportamiento reproductivo en ausencia de una expresión analítica para la función de verosimilitud. Para esto, se utilizan técnicas de Computación Bayesiana Aproximada (ABC) y se analiza la incertidumbre asociada a las predicciones del modelo. Se trabaja con el modelo *Comfert*, un modelo de microsimulación que modela las trayectorias reproductivas de una cohorte de mujeres en un régimen de fecundidad natural, es decir, en ausencia de intentos dirigidos a prevenir nacimientos. Con estas trayectorias simuladas se obtienen las tasas específicas de fecundidad por edad para dicha cohorte y se utiliza esta información para ajustar el modelo a las tasas observadas en una población histórica. Los datos utilizados provienen de una cohorte de Hutteritas, una comunidad anabaptista frecuentemente estudiada en Demografía por su rechazo del uso de métodos anticonceptivos y el nivel elevado de su fecundidad. Por otro lado, como objetivo secundario, se plantea el estudio sobre la incertidumbre en la predicción del modelo anteriormente descrito. Para ello, se realiza primero una estimación puntual considerada óptima (aquella que minimiza el Error Cuadrático Medio) y luego se construyen diferentes intervalos de credibilidad. Si bien todos se encuentran contruidos al 95 %, la incertidumbre varía notablemente en función de los niveles de tolerancia considerados en el algoritmo. Los resultados obtenidos ilustran la utilidad del enfoque bayesiano en la realización de inferencia estadística sobre modelos computacionales, al igual que una relación positiva entre el nivel de incertidumbre y el incremento en la tolerancia del algoritmo.

Metodología

El modelo con el se que trabaja, *Comfert* [2], simula trayectorias reproductivas de una cohorte de mujeres, para ajustar los parámetros del modelo se utilizan una serie de tasas específicas de fecundidad por edad de una cohorte de mujeres de la población Hutteritas, una comunidad religiosa Anabaptista con un estilo de vida tradicional que incluye el rechazo al uso de métodos anticonceptivos. A partir de esta secuencia es posible calcular los indicadores del nivel de la fecundidad más frecuentemente utilizados, como las tasas específicas de fecundidad por edad, definidas como:

$${}_n F_x [0, T] = \frac{{}_n B_x [0, T]}{{}_n L_x [0, T]}$$

siendo el numerador la cantidad de nacimientos de mujeres de edades entre x y $x + n$ sobre la población promedio (en T años) de este grupo de mujeres. Como se puede ver, esta tasa mide la cantidad promedio de hijos por mujer a la edad x .

Técnicas utilizadas

- Modelo de micro simulación
- Computación Bayesiana Aproximada (ABC)

Palabras Clave: Computación Bayesiana Aproximada, Demografía, Fecundidad, Hutteritas, Modelos Computacionales

¹e-mail: nacho.baccino.3323@gmail.com, mauroloprete1@gmail.com

²ciganda@demogr.mpg.de

Referencias

Artículos

- [2] Daniel Ciganda y Nicolas Todd. “The limits to fertility recuperation”. En: WP-2019-024 (ene. de 2019). Apéndice B. DOI: 10.4054/MPIDR-WP-2019-024. URL: <https://ideas.repec.org/p/dem/wpaper/wp-2019-024.html>.
- [3] Garrett M Dancik y Karin S Dorman. “mleqp: statistical analysis for computer models of biological systems using R”. En: *Bioinformatics* 24.17 (2008), pág. 1967.
- [5] Michael U. Gutmann y Jukka Corander. “Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models”. En: (2015). arXiv: 1501.03291 [stat.ML].
- [6] Louis Henry. “Some data on natural fertility”. En: *Eugenics Quarterly* 8.2 (1961), págs. 81-91. DOI: 10.1080/19485565.1961.9987465.
- [7] Anastasia Kostaki y Peristera Paraskevi. “Modeling fertility in modern populations”. En: *Demographic Research* 16 (2007), págs. 141-194. DOI: 10.4054/demres.2007.16.6.
- [9] Carl Schmertmann. “A system of model fertility schedules with graphically intuitive parameters”. En: *Demographic Research* 9 (2003), págs. 81-110. DOI: 10.4054/demres.2003.9.5.
- [10] Mindel C. Sheps. “An Analysis of Reproductive Patterns in an American Isolate”. En: *Population Studies* 19.1 (1965), pág. 65. DOI: 10.2307/2173165.

Paquetes de R

- [1] Rob Carnell. *lhs: Latin Hypercube Samples*. R package version 1.1.1. 2020. URL: <https://CRAN.R-project.org/package=lhs>.
- [4] Matt Dowle y Arun Srinivasan. *data.table: Extension of ‘data.frame’*. R package version 1.14.0. 2021. URL: <https://CRAN.R-project.org/package=data.table>.
- [8] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2021. URL: <https://www.R-project.org/>.
- [11] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN: 978-3-319-24277-4. URL: <https://ggplot2.tidyverse.org>.

Efecto de la posición del sensor sobre la clasificación de un fallo en las cajas de engranajes.

Antonio Pérez-Torres, Susana Barceló-Cerdá, Ana Debón, René-Vinicio Sánchez¹

Universidad Politécnica de Valencia, Universidad Politécnica Salesiana

En la industria, las cajas de engranajes utilizadas por su eficiencia en la transmisión de potencia son componentes críticos. Si se produce algún fallo o mal funcionamiento en las mismas detectarlo con suficiente antelación es crucial. Debido a esto, es necesaria la monitorización continua del estado de condición analizando las señales que se pueden obtener a través de sensores.

En este trabajo se busca determinar si el cambio en la posición de un sensor influye en la recepción de su señal y dependiendo de ello puede clasificarse el nivel de severidad de un fallo. En condiciones de laboratorio en un banco de pruebas representado en la Figura ?? se simuló la rotura del diente de un piñon con diferentes niveles de severidad en una caja de engranajes rectos, se modificaron las condiciones de operación con tres velocidades de rotación del motor mediante un variador de frecuencia y tres niveles de carga mediante un freno electromagnético; para obtener la señal de vibración se instalaron en el eje de entrada a la caja de engranajes tres acelerómetros, A1 en posición radial, A2 biaxial y A3 triaxial. En primer lugar, mediante parámetros estadísticos se calcularon indicadores de condición en el dominio del tiempo para la extracción de características contenida en la señal, y posteriormente se clasificó el nivel de severidad del fallo en el piñon mediante el método de los k vecinos más cercanos(KNN), la metodología utilizada se representa en la Figura ??.

A partir de los resultados, se concluyó que la señal del acelerómetro colocado en la posición biaxial A2 permite una clasificación más eficiente del nivel de severidad de un fallo.

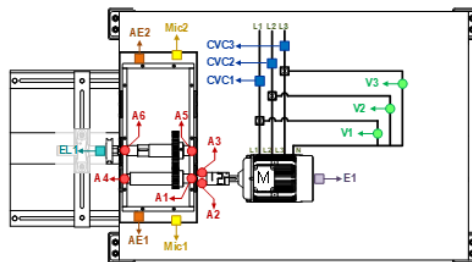


Figura 1: Banco de pruebas

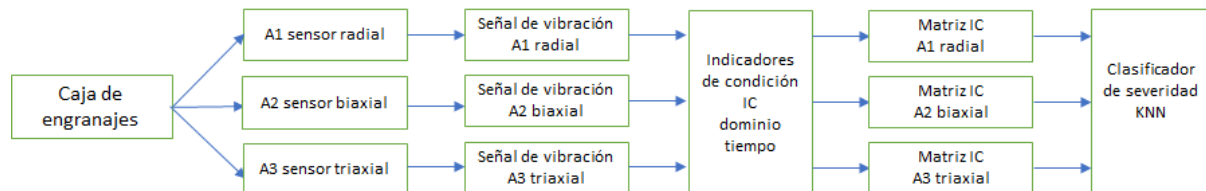


Figura 2: Metodología

Palabras Clave: Severidad de fallo, Clasificador KNN, Indicadores de condición, Vibración.

¹e-mail: jperez@ups.edu.ec, sbarcelo@eio.upv.es, andeau@eio.upv.es, rsanchez1@ups.edu.ec

Cuantificación de la incertidumbre con ecuaciones diferenciales estocásticas y aplicación a datos de pronóstico de potencia eólica

Renzo Caballero ¹, Ahmed Kebaier ², Marco Scavino ³, Raúl Tempone ^{4 5}

^{1,4} *CEMSE Division, King Abdullah University of Science and Technology (KAUST)*

² *Laboratoire de Mathématiques et Modélisation d'Evry, Univ Evry, Université Paris-Saclay*

³ *Instituto de Estadística, FCEA - Universidad de la República*

⁵ *Alexander von Humboldt Professor in Mathematics for Uncertainty Quantification, RWTH Aachen University*

En este trabajo desarrollamos una metodología orientada en datos basada en las ecuaciones diferenciales estocásticas (SDE) paramétricas de Itô que permite captar la dinámica asimétrica real de los errores de pronóstico, incluida la incertidumbre del pronóstico en el tiempo cero.

En nuestro enfoque [1], el modelo SDE contiene un término de seguimiento de la derivada del pronóstico respecto del tiempo, un parámetro de reversión a la media variable en el tiempo y un coeficiente de difusión del tipo Jacobi dependiente del estado del sistema. Caracterizamos la estructura del término de deriva del modelo basándonos en una sólida teoría matemática que asegura la existencia y unicidad de la solución fuerte de la ecuación diferencial estocástica no lineal y no homogénea en el tiempo. Un procedimiento de truncamiento regulariza la función de predicción para asegurar que las trayectorias del proceso no alcancen casi seguramente las barreras fijas en un tiempo finito.

Empleamos técnicas inferenciales basadas en métodos de verosimilitud aproximada que utilizan el procedimiento de igualdad de momentos, tanto en el espacio del error de pronóstico original como en el espacio de Lamperti, mediante procedimientos de optimización numérica. Proponemos un método de optimización de punto fijo de la verosimilitud en el espacio de Lamperti. Otro aporte novedoso es la caracterización de la incertidumbre del pronóstico en el tiempo cero que resulta crucial en la práctica. A tal efecto, caracterizamos la distribución de la densidad de transición del proceso de un instante de tiempo desconocido en el pasado al instante de tiempo en el cual se provee un pronóstico por primera vez, a través de un parámetro adicional que modela la longitud del intervalo de tiempo desconocido.

Todos los procedimientos son independientes de la tecnología de pronóstico, y permiten realizar comparaciones entre diferentes proveedores de pronósticos. Aplicamos nuestro enfoque SDE para modelar la producción eólica normalizada de Uruguay y los datos de pronóstico entre abril y diciembre de 2019.

[1] Renzo Caballero, Ahmed Kebaier, Marco Scavino, Raúl Tempone. Quantifying uncertainty with a derivative tracking SDE model and application to wind power forecast data. *Statistics and Computing* **31**, 64 (2021). <https://doi.org/10.1007/s11222-021-10040-8>

Palabras Clave: Estadística computacional, cuantificación de la incertidumbre, error de pronóstico, difusión de Jacobi no homogénea en el tiempo, espacio de Lamperti, optimización de punto fijo de la verosimilitud, potencia eólica.

¹renzo.caballerorosas@kaust.edu.sa

²kebaier@math.univ-paris13.fr

³marco.scavino@fcea.edu.uy

⁴raul.tempone@kaust.edu.sa

⁵tempone@uq.rwth-aachen.de

Convirtiendo datos en información de utilidad. Aplicación práctica empleando el Banco de datos del Centro de Investigaciones Sociológicas

Virgilio Pérez, Cristina Aybar y Jose M. Pavía¹

GIPEyOP, Universidad de Valencia

Es bien sabido que las nuevas tecnologías proporcionan cantidades ingentes de datos. No obstante, es necesario establecer procedimientos que permitan transformar estos datos en información útil. En el presente trabajo se proponen métodos y soluciones para la obtención y generación de datos de calidad, desarrollando herramientas metodológicas y conjuntos de datos provenientes de entornos naturales, que permitan encontrar respuestas a muchas de las preguntas planteadas por investigadores, ayudando a avanzar en la comprensión de cómo ha ido cambiando la sociedad con el paso del tiempo y mediante qué mecanismos la ciudadanía ha ido modificando sus escalas de valores y preferencias.

En España disponemos de una base de datos importante que puede permitir abordar esta cuestión desde una perspectiva longitudinal, observando la evolución del comportamiento de nuestros conciudadanos a lo largo de los últimos 40-50 años. Una base de datos que alberga un gran volumen de información y que, por tanto, es de enorme interés para la sociedad, además de poseer una considerable relevancia científica. Nos referimos al Banco de datos que el Centro de Investigaciones Sociológicas (CIS) pone a disposición de la comunidad científica y de la sociedad en general.

Las encuestas realizadas por el CIS son un referente de los estudios de opinión pública en España, recogiendo de manera periódica y continuada información relacionada con variables sociodemográficas, socioeconómicas y políticas, entre otras. Estos estudios constituyen la base de datos más completa y amplia de España y una de las más valiosas de cuantas se pueden encontrar en las democracias occidentales (dos mil estudios desde 1963). No obstante, esta base de datos se considera sustancialmente inexplorada, posiblemente por la dificultad que supone manejar tal volumen de datos.

Para extraer información de grandes bases de datos es necesario trabajar de forma sistemática y automatizando procesos. En el caso concreto de la información que ofrece el CIS, es preciso comenzar elaborando un mapa temático de sus estudios, utilizando técnicas de text mining para extraer las palabras clave y de procesamiento de lenguaje natural para establecer significados, con un doble objetivo. Primero, identificar cada pregunta dentro de cada barómetro y, segundo, asociar preguntas y temáticas entre barómetros. El resultado de este proceso es la creación de un diccionario que facilita moverse entre estudios y preguntas, herramienta que permite abordar/plantear cuestiones a medida.

Además, la metodología propuesta permite, por un lado, construir tablas ideológicas cohorte generacionales, utilizando técnicas de dobles ventanas móviles (edad-año) y, por otro lado, resolver/mitigar uno de los problemas que presentan las bases de datos provenientes de encuestas: el gran número de vacíos u observaciones perdidas que contienen (mediante métodos de imputación múltiple o algoritmos de machine learning). Por ello, este estudio sienta las bases para poder comparar ambos métodos de imputación, circunstancia que permitirá responder a la pregunta de cuál de las dos aproximaciones es más adecuada para imputar valores faltantes en las encuestas del CIS.

Palabras Clave: bases de datos de calidad, información útil, herramientas metodológicas, CIS.

¹e-mail: virgilio.perez@uv.es, cristina.aybar@uv.es, pavia@uv.es

Análisis de Simulación de la Gráfica de Control Poisson INAR(1) CUSUM Winsorisada

Sergio Martín Buzzi, Andrea Fabiana Righetti, Rodrigo Porta, María Virginia Bederián, Silvia Joeques¹

Instituto de Estadística y Demografía Hebe Goldenhersch, Facultad de Ciencias Económicas, Universidad Nacional de Córdoba

En este trabajo se continúan las investigaciones previas sobre el desempeño de gráficas de control de la cantidad de fallos en un contexto de alta calidad. Las intervenciones realizadas para mejorar la calidad generan que la cantidad de fallos presente mas ceros de lo habitual, por lo que una forma de modelarla es por medio de la introducción de exceso de ceros. Como la variable bajo estudio tiene como base una distribución similar a la Poisson pero potencialmente autocorrelacionada, para modelar a la misma se emplea un proceso ZINAR(1) con distribución marginal Poisson. De acuerdo a la literatura, si bien la gráfica de control Poisson INAR(1) CUSUM presenta cierta robustez ante desviaciones de la distribución Poisson, dicha robustez puede ser mejorada si las observaciones son winsorizadas antes de calcular el estadístico CUSUM. Dado que el exceso de ceros constituye un caso particular de desviación de la distribución Poisson, en este trabajo se estudia si la winsorización mejora el desempeño de la gráfica Poisson INAR(1) cuando hay exceso de ceros. Con dicha finalidad, se simulan series con distribución ZINAR(1) con distribución marginal Poisson con diversos valores del parámetro de exceso de ceros y a partir de las mismas se comparan los ARL_1 obtenidos de emplear las gráficas Poisson INAR(1) CUSUM y la Gráfica Poisson INAR(1) CUSUM Winsorizada.

Palabras Clave: Gráfica Poisson INAR(1) CUSUM Winsorizada, Control estadístico de procesos, Series de tiempo.

¹e-mail: sergio.buzzi@unc.edu.ar, righettiaf@gmail.com, portarodrigo@gmail.com, vbederian@gmail.com, silviajks@gmail.com

Paquete de R *ggcleveland*: creación de gráficos para el Análisis Exploratorio de Datos

Marcos Prunello, Gonzalo Mari¹

Escuela de Estadística, Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario

El libro *Visualizing Data* de William S. Cleveland (1993) es una pieza clásica en la literatura sobre Análisis Exploratorio de Datos (AED). Habiendo sido escrito algunas décadas atrás, su contenido sigue siendo relevante ya que propone herramientas útiles para descubrir patrones y relaciones en los datos estudiados, como así también para evaluar ajustes.

El libro menciona opciones de visualización que incluyen métodos clásicos, como gráficos de cuantiles, diagramas de cajas y bigotes (*boxplots*) o histogramas, como así también otros menos divulgados, como los gráficos de dispersión-localización (*spread-location plots*), gráficos condicionantes (*coplots*), gráficos de variabilidad de residuos y valores ajustados (*residual-fit plots*) o gráficos de Media-Diferencia de Tukey (*Tukey's MD plots*).

Este segundo grupo de técnicas de visualización requiere de cierto procesamiento de los datos y elaboración de gráficos que no forman parte de las opciones tradicionales de los software. Si bien es posible encontrar código en R para su implementación, el mismo pertenece al sistema gráfico *lattice*, cuyo uso ha pasado a un segundo plano ante la popularidad que ha tomado en la última década el entorno ofrecido por el paquete *ggplot2* (Wickham, 2016).

En este trabajo se presenta el paquete de R *ggcleveland*, que ofrece funciones para elaborar con el sistema *ggplot2* las herramientas de visualización no convencionales descritas en el libro de Cleveland, así como también algunas variantes de las mismas y conjuntos de datos de prueba. Fue diseñado como material complementario en cursos sobre AED para facilitar la programación de gráficos y actualmente está disponible para ser instalado desde CRAN (*Comprehensive R Archive Network*). La página web <https://mpru.github.io/ggcleveland> presenta ejemplos de su aplicación y la documentación de las funciones incluidas en el paquete.

Referencias:

Cleveland W. (1993). *Visualizing Data*. Hobart Press.

Wickham H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag.

Palabras Clave: software, análisis exploratorio de datos, William S. Cleveland

¹e-mail: mprunello@fcecon.unr.edu.ar, gmari@fcecon.unr.edu.ar

Tema 4: Datos faltantes

Configuración de puntos en biplots GGE con datos incompletos: comparación de dos metodologías de imputación

Angelini Julia¹, Bortolotto Eugenia, Faviere Gabriela, Cervigni Gerardo, Quagliano Marta
Centro de Estudios Fotosintéticos y Bioquímicos, CONICET - UNR

El éxito de los programas de fitomejoramiento se refleja en su capacidad para seleccionar genotipos (G) con buena adaptación y estabilidad a la región productora. Los ensayos multi-ambientales (MET) son herramientas importantes para evaluar el desempeño de cultivares en distintos ambientes (E) y analizar la interacción genotipo por ambiente (GE) que constituye una limitación en la selección de genotipos superiores. El uso del modelo de regresión por sitio (SREG) ha ganado popularidad entre los fitomejoradores como método de análisis gráfico para el análisis de datos MET. En su forma regular, SREG explora la matriz de G y GE realizando una descomposición en valores singulares (DVS) sobre la matriz residual de un ANOVA que sólo considera el efecto E. El resultado de los dos primeros términos multiplicativos de la SVD a menudo se presentan en un biplot llamado GGE. Sin embargo, SREG no siempre es lo suficientemente eficiente para analizar la estructura de datos MET de programas de mejoramiento vegetal. Tiene serias limitaciones frente a información faltante y, apesar de que los MET están diseñados para que todos los genotipos se evalúen en todos los ambientes, la presencia de valores perdidos es muy común. En este trabajo se compara la consistencia en la configuración de puntos del biplot GGE al estimar valores faltantes con los dos métodos más utilizados en este contexto: EM-GGE propuesto por Yan (2013) y EM-AMMI propuesto por Gauch (1992).

La eficiencia de los métodos se evaluó mediante un estudio de simulación basado en un conjunto de datos reales. Se contaba con el rendimiento de 18 genotipos de trigo de invierno en nueve ambientes de Ontario. Se generaron valores perdidos según el mecanismo completamente al azar (MCAR) y dos no aleatorios (MNAR), en diferentes proporciones (5, 10, 20, 30, 40 y 50 %). Los MNAR se obtuvieron eliminando un subconjunto de genotipos de un ambiente, y registros de un genotipo en diferentes ambientes. Cada escenario (combinación de mecanismos de pérdidas y proporción de datos faltantes) se repitió 10000 veces y los métodos para tratar datos faltantes fueron aplicados en cada iteración.

Dado que la importancia en el biplot GGE radica en la posición relativa de los genotipos, se evaluó la correlación entre la distancia de ellos en el subespacio generado con los datos imputados y completos. Además se calculó el ángulo formado entre los autovectores obtenidos con y sin información faltante, la variabilidad explicada por los dos primeros términos multiplicativos de la SVD y las iteraciones requeridas para la convergencia.

Diferencias mas pronunciadas entre los métodos fueron observadas cuando las pérdidas ocurren aleatoriamente en la matriz de datos. En este patrón de pérdida, ninguno de los métodos logró la convergencia para porcentajes mayores al 20 %, y la performance de ambos fue empeorando a medida que los niveles de observaciones ausentes aumentaba. El método de estimación EM-GGE fue capaz de mantener la conformación original del biplot ya que las correlaciones fueron más cercanas a 1 y ángulos a cero en las 10000 iteraciones, cuando las pérdidas eran MCAR. Sin embargo, los porcentajes de variabilidad explicado luego de imputar con EM-AMMI fueron levemente mas similares al real. Por el contrario, ambos métodos fueron consistentes en los patrones de biplot indicando una predicción exitosa en los mecanismos MNAR, aunque el EM-GGE requirió menos iteraciones para lograr la convergencia. Por lo tanto, el uso de los procedimientos de estimación de valores perdidos permitió que los datos incompletos se analizaran eficazmente mediante el biplot GGE. Como resultado, los genotipos no evaluados en el mismo ambiente y los que no han sido completamente probados pueden ser evaluados críticamente.

Palabras Clave: biplot GGE, fitomejoramiento, imputación, regresión por sitio

¹e-mail: jangelini_93@hotmail.com

Diferentes tasas de error y su efecto en las pruebas de comparaciones múltiples en ensayos de nutrición

Palacios, Luciano 1; Filippini, Susana 1; De Franceschi, Mauricio 2 ¹

1Div. Estadística, Dpto Cs Básicas, 2Avicultura, Dpto Tecnología, Universidad Nacional de Luján

Los investigadores del área Agropecuaria, Veterinaria y especialmente los dedicados a la Avicultura, que realizan ensayos en temas de Nutrición, Manejo y Medio Ambiente, encuentran un obstáculo para la toma de decisiones respecto a comparaciones de distintos grupos tratados. Los procedimientos de comparaciones múltiples permiten detectar relaciones significativas entre las medias de los tratamientos, siendo este es el objetivo de muchos experimentos con animales, (Atil, et al 2001, Casas Cardoso, et al 2008, Mendes, et al 2018). Muchas de las pruebas de comparaciones múltiples (PCM) con el objetivo de controlar la tasa de error experimental, o sea de la probabilidad de alcanzar al menos una falsa diferencia significativa, son utilizadas en los análisis de datos, tal es el caso de la prueba de Tuckey (Tuckey, 1949), Duncan (Duncan, 1955), Scheffé (Scheffé, 1953), Fisher, entre otras, siendo algunas más potentes y liberales y otras de las denominadas conservadoras. El uso de pruebas significativas es una estrategia muy arraigada, pero se han sobrevaluado las bondades de dichas pruebas al considerarlas como un indicador suficiente de la veracidad de una hipótesis, omitiendo la cuantificación de las diferencias, la consideración de las diferentes tasas de error y la potencia de la prueba. El objetivo de este trabajo es seleccionar métodos de comparaciones múltiples entre medias, considerando los diferentes tipos de error que se pueden cometer cuando se aplican en ensayos experimentales que pongan a prueba diferentes grupos. El ensayo que utilizamos en esta oportunidad, se realizó en una Granja Experimental Avícola, utilizando pollos parrilleros bajo condiciones de crianza similares a las comerciales. Se aplicó un DBCA, con cuatro tratamientos, nueve repeticiones y 10 aves/corral. Los parámetros zootécnicos medidos fueron el peso vivo y la conversión de alimento. En la formulación del alimento la única fuente de variación en las dietas fueron los niveles y origen de la enzima fitasa. Del estudio de los errores y a partir de este concepto, se desprende la definición de Tasa de error por comparación (CER: Comparisonwise Error Rate), cuyo significado es la probabilidad de que el valor de un parámetro no sea contenido dentro del intervalo de confianza, debido a que las inferencias típicas son realizadas usando un 95 % de confianza o un 5 % de nivel de significancia. En el caso de las pruebas de hipótesis, el CER de 5 % , es definido como la probabilidad de rechazar la hipótesis nula siendo esta verdadera, comúnmente conocido como error de tipo 1. Otro concepto importante es la Tasa de Error por Familia (FWE: Familywise Error Rate), que es la probabilidad de hacer al menos un falso enunciado cuando la familia de inferencias es considerada en su totalidad o conteniendo (K) elementos. Se observó, que el método de Tukey (1953), controla la tasa de Error por Familia o Error de tipo I y mejora su potencia con respecto a Bonferroni y Šidák, aunque se comportó muy conservador al momento de detectar diferencias entre las medias de las dietas. La prueba de Hochberg demostró un comportamiento similar a Tukey y prudencia al momento de detectar diferencias significativas. Las pruebas de Fisher y Duncan resultaron ser más potentes y capaces de detectar diferencias entre medias con mayor facilidad y con una alta probabilidad de cometer Error de tipo I o con una Tasa de Error por comparación mayor. La Prueba de Student, Neuman y Keuls (1972) demostró un comportamiento intermedio, ya que controla las tasas FCR y por FWE, pero posee alta probabilidad de cometer error de tipo I. Las Pruebas F y Q de Ryan, Einot, Gabriel y Welsch (REGW) son pruebas de las denominadas potentes y ejercen un buen control de la FWE. El método menos potente resultó ser Scheffé (1953), muy conservador para seleccionar medias de tratamientos, y un elevado control de la tasa de error para todo tipo de comparaciones independiente de la cantidad de comparaciones.

Palabras Clave: Avicultura - fitasas - comparaciones múltiples - tasas de error

¹email: lucianofepalacios@gmail.com, filippin@agro.uba.ar, mauriciodfp@gmail.com

Inferencia en Modelos de Regresión Multivariada con Errores de Medición

Gabriela Sandoval, Manuel Galea, Reinaldo Arellano ¹

Departamento de Estadística, Pontificia Universidad Católica de Chile

Los modelos de regresión multivariados son útiles en diferentes campos de estudio. Sin embargo, las variables predictoras podrían estar medidas con error. Los modelos de regresión multivariados con errores de medición - MRMME, por sus siglas en inglés, toman en cuenta la presencia de errores de medición en las covariables y en este trabajo se desarrolla el proceso de inferencia incluyendo estimación y test de hipótesis. Además, se considera el supuesto clásico de normalidad y para el proceso de estimación se propone seguir la metodología de máxima verosimilitud y algoritmo EM. Con la finalidad de evaluar la calidad de las estimaciones se muestran resultados de simulaciones realizadas con distintos números de covariables, componentes de la variable respuesta y observaciones. Para ilustrar la metodología se presenta una aplicación, cuyos resultados se comparan con los obtenidos al ignorar el error de medición. Adicionalmente, se posibilita el uso de este modelo a través de un paquete de R, en el que se implementa el proceso de estimación y herramientas para evaluar la bondad de ajuste.

Especificación del modelo

Sea n el número de observaciones, $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})^T$ el vector de covariables observadas de dimensión p para la observación i , $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ el vector no observado de los valores reales de las covariables para la observación i , $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iq})^T$ el vector de variables de respuesta de dimensión q para la observación i , para $i = 1, \dots, n$. El MRMME está dado por:

$$\begin{aligned} \mathbf{X}_i &= \mathbf{x}_i + \boldsymbol{\epsilon}_{i1} \\ \mathbf{Y}_i &= \mathbf{a} + \mathbf{B}\mathbf{x}_i + \boldsymbol{\epsilon}_{i2}, \end{aligned} \quad (1)$$

donde, $\mathbf{a} = (a_1, \dots, a_q)^T$ es un vector de parámetros de dimensión q , $\mathbf{B} = [(b_{ij})]$ una matriz $q \times p$ de coeficientes, $\boldsymbol{\epsilon}_{i1}$ y $\boldsymbol{\epsilon}_{i2}$ son vectores aleatorios de errores de medición de dimensión p y q respectivamente. Sean $\boldsymbol{\epsilon}_i = (\boldsymbol{\epsilon}_{i1}^T, \boldsymbol{\epsilon}_{i2}^T)^T$ y $\mathbf{Z}_i = (\mathbf{X}_i^T, \mathbf{Y}_i^T)^T$ el modelo definido en ec. (1) se puede escribir como:

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{X}_i \\ \mathbf{Y}_i \end{pmatrix} = \boldsymbol{\alpha} + \boldsymbol{\Lambda}\mathbf{x}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n, \quad (2)$$

donde $\boldsymbol{\alpha} = \begin{pmatrix} \mathbf{0} \\ \mathbf{a} \end{pmatrix}$ y $\boldsymbol{\Lambda} = \begin{pmatrix} \mathbf{I} \\ \mathbf{B} \end{pmatrix}$, y $\mathbf{0}$ es un vector de ceros de dimensión $p \times 1$ e \mathbf{I} es la matriz identidad de orden p . Luego, $\boldsymbol{\alpha}$ es un vector de $r \times 1$ y $\boldsymbol{\Lambda}$ es una matriz de $r \times p$, con $r = p + q$. Los vectores aleatorios $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_n$ son independientes con media cero y matriz de covarianza $\boldsymbol{\Sigma}_\epsilon = \text{diag}(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22})$, donde $\boldsymbol{\Sigma}_{11}$ y $\boldsymbol{\Sigma}_{22}$ son las matrices de covarianza de los vectores $\boldsymbol{\epsilon}_{i1}$ y $\boldsymbol{\epsilon}_{i2}$, respectivamente. En este caso, se asume que $\boldsymbol{\epsilon}_i$ sigue una distribución normal multivariada $N_r(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$. Para evitar problemas de identificabilidad se considera la restricción $\boldsymbol{\Sigma}_\epsilon = \phi\mathbf{I}$. En este trabajo se considera el modelo estructural, es decir, los \mathbf{x}_i son vectores aleatorios independientes e idénticamente distribuidos con distribución $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. También se considera que \mathbf{x}_i es independiente de $\boldsymbol{\epsilon}_i$, para $i = 1, \dots, n$. Por lo tanto, los vectores \mathbf{Z}_i son iid $N_r(\boldsymbol{\eta}, \boldsymbol{\Psi})$, con

$$\boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{a} + \mathbf{B}\boldsymbol{\mu} \end{pmatrix} = \boldsymbol{\alpha} + \boldsymbol{\Lambda}\boldsymbol{\mu} \quad \text{y} \quad \boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Sigma} + \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}\mathbf{B}^T \\ \mathbf{B}\boldsymbol{\Sigma} & \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T + \boldsymbol{\Sigma}_{22} \end{pmatrix} = \boldsymbol{\Lambda}\boldsymbol{\Sigma}\boldsymbol{\Lambda}^T + \boldsymbol{\Sigma}_\epsilon. \quad (3)$$

El modelo expuesto en las ec. (2)-(3) corresponde al MRMME.

Palabras Clave: measurement error model, multivariate analysis, linear regression, estructural model .

¹e-mail: fgsandoval@mat.uc.cl, mgalea@mat.uc.cl

Datos faltantes de ingresos en la EPH: una revisión del supuesto MAR

Adrián Moneta Pizarro, María Alejandra Juárez y Leiza Natalia Camilo Caro¹

*Departamento de Estadística y Matemática, Facultad de Ciencias Económicas,
Universidad Nacional de Córdoba*

Un problema frecuente cuando se trabaja con datos de la Encuesta Permanente de Hogares (EPH) del Instituto Nacional de Estadísticas y Censos (INDEC) de Argentina, es la presencia de no respuesta y valores perdidos en los ingresos monetarios. Este fenómeno afecta especialmente a quienes requieren estos datos para la clasificación de individuos y hogares con el objetivo de estimar de manera insesgada relevantes indicadores socioeconómicos para las políticas públicas tales como, por ejemplo, las tasas de pobreza e indigencia.

La literatura disponible da cuenta de un variado conjunto de métodos para lidiar con esta dificultad: desde métodos tradicionales de eliminación y reponderación, hasta técnicas más avanzadas de imputación de datos faltantes. Sin embargo, la eficacia de cada uno de estas técnicas depende críticamente del mecanismo de generación de los valores perdidos. En general, estos procedimientos parten del cumplimiento de algún supuesto sobre el mecanismo de generación de los valores faltantes y requieren, como mínimo, que los datos omitidos sean *missing at random* (MAR). Esto es que los valores perdidos aparezcan aleatoriamente condicionados a determinadas variables observadas. Por lo tanto, es de suma relevancia determinar si las personas que no declaran sus ingresos reúnen características particulares que las distinguen de los declarantes.

En este trabajo se analizan las diferencias entre declarantes y no declarantes de los ingresos monetarios en la EPH, centrandó la atención en los datos del IV trimestre 2020. Dada la naturaleza categórica de los datos observados disponibles, en primera instancia se realiza un análisis de correspondencias múltiples (ACM) y pruebas chi-cuadrado de independencia para identificar de manera exploratoria en qué variables se observan diferencias significativas. En segunda instancia y sobre la base del primer análisis, se lleva a cabo una regresión logística para determinar cuáles son los factores principales que contribuyen a explicar la falta de respuesta.

Los resultados señalan que hay diferencias significativas entre declarantes y no declarantes en un conjunto importante de características relevadas por la EPH. No obstante, también indican que el sexo, la región geográfica, la educación, el sector de actividad (público/privado) y el tipo de ocupación, son los principales determinantes de la no declaración de ingresos. De acuerdo a estos hallazgos, quienes tienen mayor propensión a no informar sus ingresos monetarios son los varones casados con educación secundaria o universitaria, que trabajan en el sector privado como cuentapropistas y residen en el área metropolitana de Buenos Aires o en la región pampeana.

Palabras Clave: datos faltantes, ingresos, análisis multivariado, Argentina

¹e-mail: amoneta@eco.uncor.edu, juarezmaale@hotmail.com, leizacamilo@gmail.com

Tema 5: Demografía

Transición a la adultez en Argentina a inicios del siglo XXI

Stanecka, Nancy Susana ¹

*Instituto de Estadística y Demografía "Dra. Hebe Goldenhersch", Facultad de Ciencias Económicas,
Universidad Nacional de Córdoba*

El pasaje de la niñez a la adultez, se manifiesta con características y duraciones diferentes de acuerdo a las distintas culturas, razas, clases sociales y género. Miradas en distintos sentidos llevan a concluir que los límites de esta etapa de la vida son cada vez más difusos. Sin embargo, la transición a la adultez puede medirse a través del análisis de algunos de los marcadores ocupacionales y demográficos que son considerados determinantes claves de esta etapa del ciclo de la vida.

Este estudio tiene por objetivo indagar sobre los cambios de patrones de transición a la vida adulta de los jóvenes de entre 15 y 29 años de la República Argentina a inicios del siglo XXI, a partir de una comparación entre los años 2001 y 2018.

En tal sentido, se analizan algunos de los eventos demográficos tales como: salida de la escolaridad, ingreso al mercado laboral, emancipación residencial y la unión conyugal, a través de variables proxy (asistencia a establecimiento educativo, condición de actividad, parentesco con el jefe y situación conyugal) presentes en las Encuestas Permanentes de Hogares. Esta encuesta constituye una fuente posiblemente limitada de información socio-demográfica y económica, pero disponible en Argentina y razonablemente actualizada.

Se sigue la metodología adoptada por Melo Vieira y Pau Miret (2010), quienes parten del índice de entropía de Theil (1972) y lo relacionan con los eventos propios de la transición a la adultez, generando la entropía de combinación de estatus para cada edad. Esto permite obtener en una medida concreta y resumida la intensidad de la transición hacia la adultez y el aporte a la misma de los distintos eventos que determinan la emancipación del joven.

Se concluye que, a partir de cierta edad, se produce un cambio significativo en la curva de entropía de un año con respecto al otro. Además, se encuentra diferencias por sexo en lo que respecta a la cúspide de máxima entropía y al patrón de transición.

Finalmente, se calculan índices de entropía reducida, obtenidos a través de la eliminación de una o más de las variables en estudio, y se compara la heterogeneidad observada con la heterogeneidad predicha para determinar si el estatus o par de estatus eliminado contribuyen o no de manera significativa a elevar la entropía.

Palabras Clave: Jóvenes, Transición, Entropía

¹e-mail: nstanecka@unc.edu.ar

Intervalos de confianza/credibilidad para tasas de mortalidad infantil con enfoque superpoblacional

Fiorella Cavalleri, Juan José Goyeneche¹

Departamento de Métodos Cuantitativos, FMed e Instituto de Estadística, FCEA, Universidad de la República, Uruguay

En Medicina es común encontrarse con tasas pequeñas que son expresadas en tanto por mil o en tanto por diez mil. En particular, la Tasa de Mortalidad Infantil (TMI) toma valores pequeños que se expresa en tanto por mil. Se disponen de datos poblacionales por lo que se emplea un enfoque superpoblacional en el sentido de Stouffer. Silva Aycaguer lo define como “un universo finito de posibilidades, formado por todos los universos finitos que pudieran haberse producido en el momento de la observación y del cual nuestra población puede considerarse una muestra aleatoria”.

Valores cercanos a cero de las tasas hacen que la aproximación por el Teorema Central del Límite no tenga propiedades deseables. Se comparan tres métodos de inferencia clásica para intervalos de confianza (Wald, inversión de prueba de hipótesis y método del score) y un método bayesiano para intervalo de credibilidad.

Palabras Clave:Bioestadística, test Binomial, intervalos de confianza, intervalo de credibilidad

¹e-mail: fcavalleri1226@gmail.com, jjgoye@iesta.edu.uy

Pronóstico de la tasa de natalidad haciendo uso del Análisis de Datos Simbólicos

Salas Andrea Mariel, Martínez Cecilia Evelyn, Mallea Lilian Adriana¹

FFHA, Departamento de Matemática, Universidad Nacional de San Juan

En general, una observación clásica es un simple valor que no posee variabilidad inherente, y por ello solamente se analiza la variación entre observaciones. En contraste, el Análisis de Datos Simbólicos (ADS) no sólo considera la variación entre los distintos valores sino también la variación interna de cada observación. Es por ello que este tipo de análisis nos permite una nueva manera de procesar información de diversas clases o grupos. Una serie temporal es el resultado de la observación de los valores de una variable a lo largo del tiempo en intervalos regulares. Las series temporales de intervalo (STI) son un tipo de series donde cada observación se describe mediante un intervalo o rango de valores. Como cada observación está representada mediante un dato simbólico, en este caso, un intervalo, lo natural es abordarlas desde la perspectiva del ADS. Un indicador como la tasa de natalidad, podemos pensarlo variando en un rango de valores, en un periodo de un año. El conocimiento de la predicción del intervalo de valor mínimo y máximo para el periodo siguiente permite pronosticar el rango en que se encontrará. En este sentido, se busca presentar una herramienta potente del ADS para el pronóstico de estadísticas vitales. Poder predecir esta tasa abre puertas a nuevos conocimientos y desarrollos de estrategias, en vistas de poder comprender fenómenos tales como el crecimiento de la población mundial y la transición demográfica.

Para este trabajo, consideramos como fuentes al Instituto Nacional de Estadísticas y Censos (INDEC) y la Dirección de Estadísticas e Información de Salud (DEIS). El método adaptado considerado para la predicción fue el Alisado Exponencial Simple (AES). El error considerado es el error cuadrático medio cometido por el método considerado en el periodo de interés (RECEM).

Los resultados fueron obtenidos por medio del paquete RSDA y time series del lenguaje de programación, Software R. Para el método de predicción adaptado se hizo uso de algoritmos desarrollados en Octave (versión libre de Matlab).

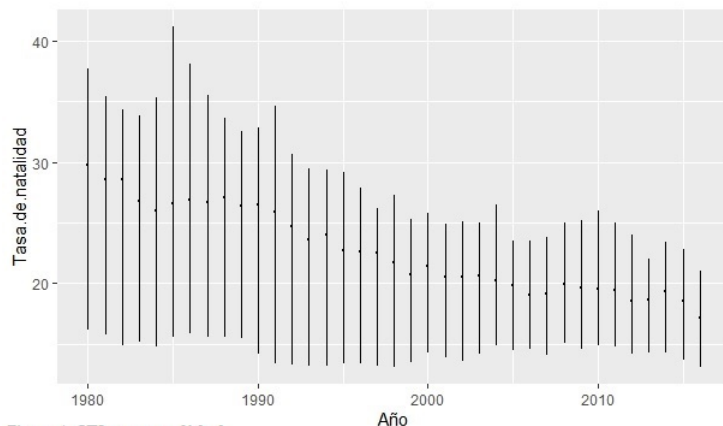


Figura 1: STI tasa natalidad

Predicción 2017		
Modelo	Mínimo	Máximo
AES (Método clásico)	13.10006	21.71175
AES (Método adaptado STI)	13,1	21,0

Tabla 1: Predicción para los valores máximos y mínimos del año 2017

RECEM		
Modelo	Mínimo	Máximo
AES (Método clásico)	1,2183	1,0824
AES (Método adaptado STI)	1,2182	0,99825

Tabla 2: Error cuadrático medio escalado

Error Porcentual		
Modelo	Mínimo	Máximo
AES (Métodos clásicos)	11.96632479%	2.8992891%
AES (Método adaptado STI)	11.96581197%	0.473933649%

Tabla 3: Error porcentual

Palabras Clave: Modelos para datos con estructura temporal, Análisis simbólico, Series temporales, Tasa de natalidad

¹e-mail: andreamarielsalas@gmail.com, ceciliaevelyn99@gmail.com, lamallea@gmail.com

Proyecciones probabilísticas de la fecundidad en Argentina

Lucía Andreozzi¹, Bruno Ribotta², María Teresa Blaconá³¹

1. Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística, Facultad de Cs. Económicas y Estadística, Universidad Nacional de Rosario - CONICET
2. Centro de Investigaciones y Estudios sobre Cultura y Sociedad, Departamento, Universidad Nacional de Córdoba - CONICET
3. Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística, Facultad de Cs. Económicas y Estadística, Universidad Nacional de Rosario

Resumen

En los últimos años se ha propuesto una importante cantidad de métodos estadísticos demográficos. La gran mayoría han sido desarrollados con la finalidad de pronosticar las componentes demográficas y/o medidas derivadas a partir de la suposición de un modelo subyacente. El presente trabajo pretende realizar un ejercicio comparativo integral a través la estimación y pronóstico de la fecundidad a partir de tres propuestas: Métodos clásicos de pronóstico tales como los Modelos ARIMA y los Suavizados Exponenciales, Modelos para Datos Funcionales (MDF) y Modelos jerárquicos Bayesianos (MJB), constituyendo un primer paso hacia el estudio de las proyecciones de población derivadas de cada una de ellas empleando datos de la Argentina.

Como primer resultado se presentan las estimaciones obtenidas para la Tasa Global de Fecundidad desde 2015 a 2095 (en períodos quinquenales), mediante los distintos enfoques y el análisis detallado de los resultados en relación al modelo del cuál derivan, los datos base y la complejidad de los mismos. El ejercicio que se tiene como horizonte final es la estimación de la mortalidad y la fecundidad a través de los tres métodos mencionados; ARIMA/Suavizados, MDF, MJB, para luego integrarlos en proyecciones de población.

Palabras Clave: Modelos para Datos Funcionales, Modelos jerárquicos Bayesianos, Tasa Global de Fecundidad, Argentina, Proyecciones Demográficas.

¹e-mail: andreozzi.lu@gmail.com, bruno.ribotta@gmail.com, mblacona@fcecon.unr.edu.ar

DETERMINANTES DE LA FECUNDIDAD. UN ANÁLISIS ESPACIAL EN LA PROVINCIA DE CÓRDOBA

David Giuliodori ¹, Mariana Gonzalez²

Facultad de Ciencias Económicas, Departamento, de Estadística y Matemática, Universidad Nacional de Córdoba

La fecundidad, como otros fenómenos demográficos, posee particularidades sociales, culturales y económicas de índole local que deben ser consideradas en el análisis. La mayor parte de la bibliografía, sin embargo, ha enfatizado el estudio de la fecundidad a nivel individual o a nivel agregado, utilizando datos que no permiten capturar el impacto de las interacciones sociales en el espacio.

En estudios anteriores hemos detectado, mediante métodos exploratorios, que existen evidencias de patrones espaciales vinculados a las principales variables socioeconómicas que afectan al nivel de fecundidad en la Provincia de Córdoba, trabajando con datos provenientes principalmente del Censo 2010, además de otras fuentes de información secundarias. Verificada la dependencia espacial, se aplicó la técnica de regresiones geográficamente ponderadas, como extensión de la estimación de mínimos cuadrados ordinarios al caso de datos georeferenciados, ajustando un modelo que permite el análisis local de la regresión lineal. En lo referente al impacto de las variables socioeconómicas, se encontró que el nivel educativo y la situación de pareja de la mujer, son los principales factores que explican el nivel de fecundidad. Otras variables incluidas en el análisis, en cambio, mostraron un efecto homogéneo a través del espacio.

En este trabajo se avanza sobre otras estrategias de modelación que permitan incorporar elementos espaciales en el error y en la ecuación principal simultáneamente, como el modelo SARAR y el modelo espacial de Durbin.

Palabras Clave: fecundidad, Provincia de Córdoba, GWR, heterogeneidad espacial

¹e-mail: dgiuliodori@unc.edu.ar

²e-mail: mariana.gonzalez@unc.edu.ar

Análisis de la mortalidad en la Provincia de Tucumán (Argentina) durante los años 2016 – 2018

Maira Celeste Lescano.¹

Argentina, Tucuman, Universidad Nacional de Tucuman

Introducción: El análisis sobre mortalidad y movimiento constituyen un eje fundamental en el estudio de la dinámica demográfica de una población. **Objetivo:** Detectar las principales causas de defunción y analizar las diferentes tasas de mortalidad en los años (2016- 2018). **Materiales y métodos** Se trata de un estudio descriptivo de corte longitudinal, mediante el cual se analizó la mortalidad. **Instrumentos y técnicas** de recolección de datos La recolección de datos se hizo a través de la base de datos de la Dirección de Estadística e Información de Salud y del Instituto Nacional de Estadística y Censos (INDEC). **Procesamiento y análisis de datos:** Se realizó un análisis descriptivo de las variables en estudio. **Resultados:** La tasa de mortalidad disminuyó en el año 2018 y fue del 7,67 por cada 1.000 habitantes ,significamente su tendencia disminuye. En el año 2016 Tucumán tuvo la tasa de mortalidad más alta con 7.80 por c/ 1000 habitantes, en el año 2017, la tasa de mortalidad fue del 7,65 por c/ 1000 habitantes. La tasa de mortalidad del varón en el año 2018 fue del 8,5, lo que aumento con respecto al año 2016 que fue del 8.4, en la mujer, su tasa de mortalidad en el año 2018 disminuyó 6,9. En todos los grupos de edad se observa un notable descenso de la Tasa de mortalidad a lo largo de los últimos 3 años (2016 – 2018) pero éste es aún más marcado en los mayores de 75 años, el comportamiento de la Tasa de mortalidad por edad es homogéneo para los diferentes años; es decir que es elevada en los mayores de 75 años y descienden en menores de 5 años y jóvenes. En el año 2017 tiene la tasa de mortalidad infantil más alta con 11,55, con lo cual aumento con respecto al año 2016 con una tasa de mortalidad del 10,95. En el año 2018 la tasa de mortalidad infantil fue del 11,18 disminuyó con respecto al año anterior la tendencia de la mortalidad infantil en general aumenta a medida que pasa los años. Las principales causas detectadas de defunción en Tucumán en los años (2016-2018) son (enfermedades infecciosas y parasitarias, tumores, enfermedades del sistema circulatorio, enfermedades del sistema respiratorio), entre la edad media aproximadamente detectadas de las causas son 63 y 64 años, en la edad de 75 años en ambos sexos tiene mayores causas de defunción. **Conclusión:** el análisis de mortalidad se evidenció en la Tasa de mortalidad que desciende notablemente, se observó que un 20 por ciento de las causas principales representan un 80 por ciento de las causas, por lo tanto, centrándose en solo esas 4 causas principales se reduciría el número de causas de defunción.

Palabras Clave: tasa de mortalidad, causa de defunción , sexo, edad

¹e-mail: may_celeste2011@live.com.ar

Años Potenciales de Vida Perdidos por Siniestros de Tránsito en Uruguay

Gonzalo De-Armas, Mauro Loprete, Ramón Álvarez-Vaz¹

Instituto de Estadística, Departamento de Métodos Cuantitativos, Facultad de Ciencias Económicas y de Administración, Universidad de la República, Montevideo, Uruguay

Resumen

Los Años Potenciales de Vida Perdidos (APVP) son un indicador demográfico que permite cuantificar cuantos años, en promedio, se han dejado de vivir por un conjunto de causas de muerte, en particular. Se calculan como la sumatoria, para todos los fallecimientos, de la diferencia de años entre la edad del fallecido y una cierta edad límite L , que en la presente investigación se especifica como la esperanza de vida a la edad x . Los siniestros de tránsito en el Uruguay son una de las causas externas con mayor cantidad de decesos y una de las primeras en el grupo de edad que comprende los 15 a 45 años. Los resultados obtenidos se visualizan mediante una interfaz gráfica desarrollada mediante el paquete *shiny* en R^2 . Puede concluirse que las muertes y APVP por siniestros de tránsito tienen un perfil masculino, joven, donde las muertes entre conductores de motocicletas dominan en el aporte al total de APVP. Si se compara entre los distintos escenarios de siniestros, las muertes entre quienes viajaban en una motocicleta tienen una mediana de edad inferior a los 30 años, mientras que las muertes entre los peatones se caracterizan por una mediana superior a los 60 años.

Metodología

En el contexto de la primera transición demográfica, se ha registrado una disminución sostenida de la mortalidad, que frecuentemente se desagregan en: transmisibles, no transmisibles y externas.

En los países donde la mortalidad, ha llegado a niveles mínimos, las causas externas son una causa de muerte de alta prevalencia en los jóvenes, dentro de las causas externas, se pueden caracterizar como según [1]:

- Intencionales
- No intencionales
- Intención indeterminada

Los siniestros de tránsito, por lo tanto, al ser una causa de muerte externa y no intencional, tienen un mayor potencial para disminuir su valor con políticas públicas apropiadas.

Con motivo entonces de poder analizar el problema descrito, se plantea la pregunta: Cuál ha sido la evolución de los años potenciales de vida perdidos por mortalidad asociada a siniestros de tránsito en Uruguay entre 2012-2018? Para contestar esta pregunta se plantea como objetivo el calcular los APVP generados por siniestros de tránsito en el Uruguay, para el período 2013-2018, desagregando por el rol del fallecido (peatón, conductor o acompañante) y el tipo de vehículo. Para cumplir con este objetivo, se utilizan los datos abiertos proporcionados por UNASEV (Unidad Nacional de Seguridad Vial 2013-2019)

³

Palabras Clave: APVP, Shiny, Siniestros de Tránsito.

¹e-mail: gonzalo.dearmas@iesta.edu.uy, mauroloprete1@gmail.com, ramon@iesta.edu.uy

²Repositorio disponible aquí <https://gitlab.com/iesta.fcea.udelar/apvp-de-transitos-en-uruguay>

³Disponibles en : https://catalogodatos.gub.uy/dataset/unasev-fallecidos_siniestros_transito

Referencias

- [1] Arriaga, E. Los años de vida perdidos: su utilización para medir el nivel y cambio de mortalidad, 1996.
- [2] Baio, Gianluca. Bayesian Methods in Health Economics. Champman and Hall CRC - Biostatistics Series. 2013.
- [3] Dávila-Cervantes, C. y Pardo, A. Análisis de la tendencia e impacto de la mortalidad por causas externas: México, 2000-2013. Salud Colectiva, 12:251, 2016.
- [4] Dempsey, M. Decline in tuberculosis: the death rate fails to tell the entire storyd, 1947.
- [5] Frenk, J. Bobadilla, J. Stern, C. Frejka, T. Lozano, R. Elements for a theory of transition in health. Salud pública de México, 33:448–62, 1991.
- [6] Murray, C. J. L., Lopez, A. D., Organization, W. H., Bank, W., y of Public Health, H. S. The global burden of disease : a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020 :summary and edited by christopher j. l. murray, alan d. lopez, 1996.
- [7] Omran, A The epidemiologic transition: a theory of the epidemiology of population changes. The Milbank Memorial Fund Quarterly, vol. 49, No 4, 1971
- [8] OPS. La salud en las Américas, 1998.
- [9] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, 2020.
- [10] Salomon, J. y Murray, C. The epidemiologic transition revisited: Compositional models for causes of death by age and sex. Population and Development Review, 28:205–228, 2002.
- [11] Rstudio, Inc. Easy web application in R. 2013.
- [12] UNASEV. Fallecidos por siniestros de tránsito, datos abiertos, 2019.

Tema 6: Deporte

Estudio Sobre la Competencia Interna y el Desempeño Internacional. Un Análisis desde el Fútbol Sudamericano

Luciana Cantera, Leandro Zipitria¹

Departamento de Economía, Facultad de Ciencias Sociales, Universidad de la República, Uruguay

Este trabajo estudia la relación entre la intensidad competitiva de un mercado y el resultado que obtienen los agentes (equipos deportivos) a nivel internacional. En particular se toma el fútbol a nivel sudamericano y se evalúa el balance competitivo de las ligas nacionales que participan de la Copa Libertadores de América. De esta forma se analiza el impacto que podría tener un balance competitivo más equilibrado en las ligas nacionales sobre la probabilidad de obtener el campeonato a nivel internacional. El término balance competitivo fue introducido por Rottenberg (1956) y analizado posteriormente en múltiples investigaciones (Neale, 1964; Hodiri and Quirk, 1971 and Sloane, 1971). Dicho término establece que para maximizar el atractivo comercial (un mayor número de espectadores) de una liga deportiva, ésta debe estar compuesta por equipos de competitividad similar. Es decir, una liga esta balanceada desde el punto de vista competitivo si, a priori, todos los equipos tienen la misma probabilidad de ganar el campeonato.

Se analiza el mercado de las ligas deportivas, en especial de fútbol. Los datos deportivos presentan diversas ventajas para efectuar análisis empíricos de las teorías e hipótesis económicas.

El estudio considera 20 ediciones de la Copa Libertadores, desde el año 2000 hasta el 2019. Los países que integran el análisis son: Argentina, Bolivia, Brasil, Chile, Colombia, Ecuador, México, Paraguay, Perú, Uruguay y Venezuela. Se elaboró una base de datos que contiene las posiciones de todos los equipos en los campeonatos nacionales entre los años 2000 y 2019 para cada uno de los países considerados en el análisis y para la Copa Libertadores. Los datos utilizados fueron tomados de Wikipedia.

Se elaboró una base de datos que contiene las posiciones de todos los equipos en los campeonatos nacionales entre los años 2000 y 2019 para cada uno de los países considerados en el análisis y para la Copa Libertadores. Los datos utilizados fueron tomados de Wikipedia.

Entre los años 2000 y 2019 solo equipos de cinco países de los once considerados en el estudio fueron campeones de la Copa Libertadores. Argentina y Brasil son los dos países que obtuvieron más copas en esos años, con 8 campeonatos cada uno. Colombia obtuvo dos copas y tanto Ecuador como Paraguay obtuvieron una copa.

Al estimar el efecto del nivel de competencia en las ligas domésticas sobre el desempeño de los equipos a nivel internacional, las características no observadas de los países en relación al deporte, como la cultura, la tradición deportiva y futbolística o la organización de las ligas en niveles formativos, podrían introducir sesgos en las estimaciones de MCO en la variable seleccionada para medir el balance competitivo. Por lo que la metodología utilizada en este estudio abordar el sesgo de los estimadores generado por la heterogeneidad inobservable y aprovecha los datos longitudinales, para esto se estima un modelo de panel de efectos fijos.

Los resultados obtenidos indican que mejorar el nivel de competencia interno, o en otras palabras, tener estructuras de mercado más competitivas, con un mayor número de firmas y con un poder de mercado equilibrado impactaría positivamente sobre los resultados que obtienen algunas firmas que compiten representando al país a nivel internacional. También se encuentra que dicho efecto no se mantiene al considerar el desempeño promedio de todas las firmas que compiten a nivel internacional. Los resultados internacionales mejoran con niveles de competencia interna más altos pero dicho efecto no es el mismo para todos los equipos.

Palabras Clave: Balance Competitivo, Fútbol Sudamericano, Desempeño internacional.

¹e-mail: luciana.cantera@cienciassociales.edu.uy, leandro.zipitria@cienciassociales.edu.uy

Is football purely stochastic, made out of luck, or maybe predictable? How does Bayesian reasoning assess sports?

Leonardo Barrios Blanco¹, Diego C. Nascimento²

Departamento de Matemática, Universidad de Atacama (UDA), Chile.

Probabilistic models are useful in describing a stochastic phenomenon, particularly in the gambling context. This work analyzed the predictability of the 2020 Chilean Premier League teams considering a Bayesian hierarchical structure. Whereas a discrete (Poisson) regression structure was adopted, in order to explain the expected number of goals, of each match, as a combination of each team attack and defense power, as well as a home playing advantage parameter, thus, considering that the championship games are independent. This statistical model made it possible to quantify the defense and offense of each team, and also to group/verify the performance of the 2020 Chilean Premier League. For the estimation process, the last 5 rounds of the league were reserved only for prediction. Results showed that the major difference across the teams was related to their attack performance, which contributes to the match prediction score, therefore on the expected rank position. The Chilean Premier League, in general, showed low defense performance, that is, not been statistically meaningful, and we have also found that playing at home or away did not influence the expected number of goals. Our model predicted the 2020 Chilean Premier League accurately the TOP 5 positions, and from the positions 6-11th were well predicted (switching some positions), thus, several teams that were very close to the predicted position in which they finished fairly close to the final season result. This class of models showed to be very competitive towards football prediction and, now, validated for the Chilean Premier League, especially for championships where repeated matches are designed (that is, teams play all rounds and are not eliminated during the tournament as in a Cup).

Palabras Clave: Hierarchical Bayesian Model, Poisson Regression, Football Game prediction, Chilean premier league.

¹leonardo.barrios.2020@alumnos.uda.cl

²diego.nascimento@uda.cl

Efecto de tecnología en el Fútbol: el caso de la liga inglesa

Ignacio Alvarez-Castro, Maximiliano Machado, Andrés Sosa¹

Instituto de Estadística, Facultad de Ciencias Económicas y Administración, Universidad de la república

El fútbol es uno de los principales deportes a nivel mundial. Recientemente, un cambio importante en el juego ha sido la incorporación de *Video Assisted Referee* (VAR) para reducir errores y ganar en la justicia de las desciciones arbitrales en los partidos.

En los últimos años aparecen varios trabajos que reportan evidencia de efectos atribuidos al uso de VAR en las ligas de Italia y Alemania (Lago Peñas et al., 2019), la liga China (Bo Han, 2020) y la liga española (Errekagorri et al., 2020). Sin embargo, estos trabajos estudian diferencias observadas en el juego antes y despues de la implementación del VAR sin una estrategia de identificación que permita separar el efecto VAR de otros aspectos que cambien de una temporada a la siguiente.

Este trabajo propone incorporar una estrategia de identificación clara para investigar los efectos de la implemtación del VAR en algunas dimensiones del juego en las que es utilizado (cantidad de goles anotados, tarjetas rojas, penales). La incorporación de VAR se ha dado en forma gradual, es posible aprovechar esto para identificar el efecto que puede tener su incorporación.

La fuente principal de datos es el sitio football-data.co.uk de donde se obtienen datos de cada juego disputado en las ligas Inglesas de primer y segunda división para varias temporadas. La información disponible tiene goles convertidos, tarjetas, indicadores defensivos y de ataques, cantidad de faltas por equipo, etc. Estos datos son complementados con los penales cobrados por juego para cada equipo.

La liga inglesa de primera división, *Premier League*, incorpora VAR desde la temporada 2019-20 mientras que la segunda división, *Championship*, no utiliza esta tecnología. Se utiliza un diseño de diferencias-en-diferencias para estimar el efecto de VAR en la *Premier League*. Con especial interés en estudiar si la ventaja de localía ha sido modificada por esta tecnología. Una de las hipótesis iniciales es que el uso de VAR tiene el potencial de resultar en decisiones mas justas en el juego. Uno de los aspectos donde esto puede ser observable es la cantidad de penales al equipo visitante o equipos relativamente débiles en la liga.

Los resultados primarios indican que si bien hay diferencias observables al comparar las temporadas pre y post inclusión del VAR cuando se controlan por aspectos relevantes del juego y calidad de jueces, el efecto del VAR no parece relevante.

Palabras Clave: Inferencia Causal, Fútbol, Efecto VAR, Diferencias en Diferencias

¹e-mail: nachalca@fcea.edu.uy, andres.sosa@fcea.edu.uy

Tema 7: Econometría

La volatilidad en un mercado bursátil pequeño

Juan Carlos Abril, María de las Mercedes Abril¹

Facultad de Ciencias Económicas, Universidad Nacional de Tucumán y Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina

Numerosas series de tiempo económicas no tienen una media constante y en situaciones prácticas, frecuentemente vemos que la varianza del error observacional, condicional al conocimiento pasado, está sujeta a una sustancial variabilidad a través del tiempo. Ese fenómeno es conocido como *volatilidad*.

Existe una variedad muy grande de modelos no lineales en la literatura, útiles para el análisis de series de tiempo económicas con volatilidad. Una clase importante de ellos son 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.

Los modelos de la familia ARCH o GARCH suponen que la varianza condicional (volatilidad) depende de las observaciones pasadas. En otras palabras, si σ_t^2 es la volatilidad, la familia ARCH-GARCH supone que la misma depende de la serie y_j para $j < t$. Por otra parte, el *modelo para volatilidad estocástica* o MVE, propuesto por primera vez por Taylor (1980, 1986) no parte de este supuesto. Este modelo tiene como premisa el hecho de que la volatilidad σ_t^2 depende de sus valores pasados (σ_j^2 para $j < t$) pero es independiente del pasado de la serie bajo análisis (y_j para $j < t$).

Como aplicación, se analiza la serie del índice Merval, que es un índice bursátil que se calcula en la Bolsa de Comercio de Buenos Aires (BCBA), Argentina, desde el 30 de Junio de 1986. Mide el volumen negociado de las principales acciones que cotizan en esa bolsa.

De acuerdo a un estudio, el valor promedio de las empresas que cotizan en la BCBA es de 312 millones de dólares, cifra que ubica a Argentina en el puesto 30 entre los países que tienen mercados accionarios. Por eso es que decimos que estamos ante un mercado bursátil pequeño.

El período analizado va desde el 13 de Enero de 2003 hasta el 22 de Mayo de 2015. Son 3006 observaciones. Cubre un período en el que no hubo cambio de filiación política del gobierno. Esto elimina los efectos que pudieran haberse introducidos en el mercado por cambios en el grupo gobernante.

Para hacer el análisis usamos dos enfoques: uno basado en los modelos del tipo ARCH-GARCH, y otro basados en los modelos de volatilidad estocástica. Luego realizamos comparaciones entre estos dos enfoques.

Palabras Clave: Volatilidad, Modelos ARCH-GARCH, Modelos de espacio de estado, Filtro de Kalman, Volatilidad estocástica, Índice Merval.

¹e-mail: jabril@herrera.unt.edu.ar, mabrilblanco@hotmail.com

Curva de Beveridge (1980-2018)

Federico Molina¹

Instituto de Estadística (IESTA), Departamento de Métodos Cuantitativos, Universidad de la República

La curva de Beveridge (CB) se plantea como una relación negativa entre la tasa de vacantes laborales y la tasa de desempleo. Las vacantes se entienden como aquella posición dentro de la firma que el empleador busca llenar activamente en un periodo a determinar. Es utilizada como un proxy de la demanda laboral cuantificándose mediante los avisos laborales publicados en la prensa o portales de Internet. Los desocupados se definen como personas buscando trabajo remunerado activamente que no logran obtenerlo durante un periodo de referencia. La CB es el marco teórico dominante en el análisis macroeconómico del mercado laboral puesto que contiene información esencial para analizar los shocks, la coexistencia de vacantes y desempleo, el proceso de matching y la eficiencia.

Pese a su relevancia no es posible calcular la CB en Uruguay porque no se cuenta con información de vacantes laborales. Por ello, el primer objetivo específico de este trabajo es sistematizar datos existentes respecto a vacantes laborales cuya metodología permita una armonización de las series de vacantes. Extender el periodo de análisis en base a información recabada a partir del diario El País, clasificados laborales, *Gallito Luis* y los portales laborales *Buscojobs* y *Computrabajo* mediante scraping web y conteo de avisos en prensa laboral, imputación de datos faltantes y análisis de texto (*text mining*). Creando un índice de vacantes laborales de frecuencia trimestral de carácter público y prolongado que se convierta en una herramienta de análisis macroeconómico. El mismo siempre es condicional al departamento de Montevideo y cubre el periodo de 1980-2018. Actualmente se busca expandir hasta el año 2021 para considerar el efecto de pandemia.

Posteriormente se busca responder si existe algún cambio de pendiente o traslado en la CB en el periodo considerado, para lo cual se utilizan dos estrategias empíricas. Primero, se pone a prueba la hipótesis nula de que no existe un cambio estructural en la relación de vacantes y desempleo bajo diferentes test de quiebre estructural, se plantean test de tipo dating que obtienen la cantidad de quiebres estructurales optimizando una función objetivo bajo el algoritmo de programación dinámica y se estiman las CB para los periodos obtenidos. Finalmente, se modeliza la curva de Beveridge bajo vectores autorregresivos de parámetros variables y volatilidad estocástica para analizar cambios de la relación entre vacantes y desempleo. La identificación se realiza bajo un modelo básico de búsqueda y emparejamiento con restricciones de identificación de Cholesky y se estima de forma bayesiana mediante el algoritmo de muestreo de Gibbs (*Gibbs sampling*).

Palabras Clave: Curva de Beveridge, Vacantes laborales, TVP-VAR-SV, Quiebres estructurales

¹e-mail: federicoandresmolina@mail.com, federico.molina@fcea.edu.uy

On the dynamics of relative prices and the relationship with inflation: an empirical approach

Gabriel Brida¹, Emiliano Álvarez², Pablo Mones³

GIDE, Departamento de Métodos Cuantitativos, Facultad de Ciencias Económicas y Administración

En este trabajo se representa a la estructura de precios de la economía uruguaya como una red compleja para diferentes períodos y se introduce un conjunto de sets mínimos dominantes de precios para estudiar la estructura y evolución de los precios relativos en la economía. Debido a la naturaleza dinámica de los vínculos entre los precios, se propone también la construcción de un índice sintético dinámico, y se estudian sus propiedades y componentes.

Las propiedades estadísticas de los sets dominantes fueron estudiadas. En particular, se detecta que la frecuencia con la cual los nodos pertenecen al set dominante sigue una distribución de leyes de potencias, algo característico de los sistemas complejos.

Los resultados empíricos permiten identificar los nodos más importantes en estas redes. Con este conjunto de precios se generó un índice sintético que permite realizar proyecciones más correlacionadas con el índice original y con menor error que los índices del mismo tamaño utilizados como referencia. Los alcances de estas mejoras en las proyecciones y sus consecuencias en el diseño de políticas antiinflacionarias son también analizados. Asimismo, se constata que los precios de los alimentos son especialmente relevantes en este conjunto de precios mínimos; este resultado se analiza en el contexto de las economías en desarrollo y las industrias globalizadas. En particular, se discute si es eficiente focalizar el objetivo de inflación sobre el índice subyacente dada la influencia y el efecto contagio (o derrame) que tienen los precios de los alimentos sobre el resto de los bienes y servicios de la economía.

Palabras Clave: Análisis de redes; Set dominante; Precios relativos; Inflación

¹e-mail: gbrida@ccee.edu.uy

²e-mail: ealvarez@ccee.edu.uy

³e-mail: pmones@iesta.edu.uy

Aplicación de un modelo de ecuaciones estructurales en medición de Brand Equity entre estudiantes de la Universidad Nacional de Córdoba

Casini Rosanna ¹

Instituto de Estadística y Demografía, Facultad de Ciencias Económicas, Universidad Nacional de Córdoba

El valor de una marca es el rendimiento extra que obtienen las empresas a partir de que sus productos sean identificados con dicha marca, es un elemento diferenciador que se construye a través de las percepciones de los consumidores. Es importante que las empresas midan sus marcas con la intención de tomar mejores decisiones en relación a estos activos intangibles. El presente trabajo consiste en una investigación de tipo causal en el cual se plantean dos modelos teóricos referentes al valor de marca, sus antecedentes y consecuentes. Para ello se aplica un instrumento de medición en estudiantes de la Universidad Nacional de Córdoba. Se establecen dos poblaciones, dos categorías de producto y dos marcas respectivamente. Al trabajar con múltiples variables latentes, se recurre al empleo de ecuaciones estructurales como el método más adecuado para el análisis. Las principales conclusiones demuestran, por un lado, la existencia de un proceso de creación del valor de marca existente en la mente del consumidor y por el otro, que el valor de marca actúa como antecedente de la intención de compra y la disposición al pago de un precio primado. Debido a que la investigación fue llevada a cabo en un contexto limitado no es óptimo extrapolar dichos resultados al mercado en general. Medir el valor de marca es útil para la gestión empresarial debido a que el mismo presenta formas de diferenciación en el mercado no basadas en el precio. Los gerentes deben ser conscientes de que el valor de marca está condicionado por diferentes componentes y deben comprender qué dimensión contribuye en mayor medida a la creación del valor y así llevar adelante una mejor asignación de recursos.

Objetivos
Objetivo general Determinar cómo valoran las marcas los estudiantes de la Universidad Nacional de Córdoba pertenecientes a distintas carreras de grado.
Objetivos específicos
 • Comprobar la existencia de relaciones causales (notoriedad, calidad percibida, asociaciones de marca y lealtad) en la formación del concepto valor de marca, generando un modelo de antecedentes del constructo.
 • Validar un modelo de consecuentes del valor de marca que incluya la intención de compra y la disposición de pago a un precio primado.

Resultados
 Tabla 1: Resultados de la moderación del modelo de ecuaciones estructurales de antecedentes
 Coeficientes (Valor t) Relación Lic. en Trabajo social Lic. en Administración Lealtad –¿Valor de marca 0,8188 (22,78) 0,7140 (15,91) Asociaciones –¿Lealtad 0,5569 (6,64) 0,2813 (3,71) Calidad percibida –¿Lealtad 0,2435 (2,73) 0,4495 (6,45) Notoriedad –¿Calidad Percibida 0,4035 (5,57) 0,2861 (3,95) Notoriedad –¿Asociaciones 0,2281 (2,81) 0,2707 (3,65) Bondad de ajuste del modelo: Lic. En Trabajo Social Chi cuadrado/grados de libertad: 1,61 CFI: 0,966 TLI: 0,958 SRMR: 0,047 RMSEA: 0,057 Bondad de ajuste del modelo: Lic. En Administración Chi cuadrado/grados de libertad: 1,417 CFI: 0,978 TLI: 0,974 SRMR: 0,055 RMSEA: 0,044 Fuente: elaboración propia
 • Tabla 2: Resultados de las relaciones del modelo estructural de consecuentes
 Relación Coeficientes (Valor t) Valor de marca –¿Precio Primado 0,7311 (20,71) Valor de marca –¿Intención de compra 0,6775 (19,50) Fuente: elaboración propia

Referencias

Aaker, D., 1991. A.(1991). Managing brand equity. Simon Schuster, New York. Aaker, D.A., (1996). "Measuring brand equity across products and markets". California management review, Vol. 38 No. 3, pp 35-37. Angulo, J. R., Oliva, E. J. D., Romero, C. A. R. (2013). "Lealtad de marca: antecedentes y perspectivas de investigación". Universidad Empresa, Vol. 15, No. 24, pp. 141-163. Bagozzi, R. P., Yi, Y. (1998). "On the evaluation of structure equation models". Acad. Mark. Sci, Vol.16, No 1, pp. 76-94. Buil, I., Martínez, E., De Chernatony, L. (2010). "Medición del valor de marca desde un enfoque formativo". Cuadernos de gestión, Vol. 10, pp. 167-196. Cupani, M. (2012). "Análisis de Ecuaciones Estructurales: conceptos, etapas de desarrollo y un ejemplo de aplicación". Revista tesis, Vol. 2, pp. 186-199. MacCallum, R. C., Browne, M. W., Sugawara, H. M. (1996). "Power analysis and determination of sample size for covariance structure modeling". Psychological methods, Vol. 1, No. 2, pp. 130. Mulaik,

¹e-mail: rcasini1@gmail.com

SA, James, LR, Van Alstine, J., Bennett, N., Lind, S. y Stilwell, CD (1989). "Evaluación de índices de bondad de ajuste para modelos de ecuaciones estructurales". Boletín psicológico, Vol. 105 No. 3, pp. 430. Villarejo Ramos, Á. F. (2002). "Modelos multidimensionales para la medición del valor de marca". Investigaciones Europeas en Dirección y Economía de la Empresa, Vol. 8 No. 3, pp. 13-44.

Palabras Clave: Valor de Marca, modelo de ecuaciones estructurales antecedente y consecuente .

Propuesta de índice multidimensional de calidad del empleo para Argentina

Eduardo Bologna (a)¹, Pablo Sebastian Gomez (b)², Maximiliano Iglesias (c)³

(a) Centro de Estudios Avanzados, Facultad de Ciencias Sociales, Universidad Nacional de Córdoba

(b) CONICET, Universidad Nacional de Córdoba, Universidad Siglo 21

(c) Instituto de Estadística y Demografía, Facultad de Ciencias Económicas, Universidad Nacional de Córdoba

En este trabajo realizamos una propuesta de operacionalización del concepto de calidad del empleo a través de encuestas de operativos estadísticos oficiales y analizamos el comportamiento en tres dimensiones: el género, el estatus migratorio y la escala subnacional. Para reducir las limitaciones de un concepto polisémico, el índice elaborado toma como punto de referencia la definición propuesta por la Organización Internacional del Trabajo y se basa en aspectos objetivos de la inserción laboral. En esta instancia, se excluyen las evaluaciones subjetivas de la calidad del empleo medidas en encuestas específicas. En este aspecto, se consideran tres dimensiones: a) seguridad ofrecida por el puesto de trabajo, 2) relación entre la educación y la calificación demandada por el trabajo y c) ingresos relativos. La seguridad en el empleo se obtiene a través de la combinación de la estabilidad económica y la disponibilidad de cobertura médica. La consistencia a partir de la relación entre educación obtenida y calificación requerida por el puesto laboral. Finalmente, el ingreso se considera en dos dimensiones: 1) las horas trabajadas y 2) su importancia relativa según escalas subregionales. La fuente de datos utilizada corresponde al programa Encuesta Permanente de Hogares (EPH) del Instituto Nacional de Estadísticas y Censos de Argentina. Si bien existe abundante literatura sobre la construcción de índices multidimensionales, uno de los aspectos donde no hay consenso en la literatura está en las formas de agregar las variables y dimensiones presentes (Albo et al., 2017; Poirier et al., 2020; S Greco, 2018; Schlossarek et al., 2019). En efecto, en la creación de indicadores sintéticos multidimensionales, los ponderadores de las diferentes dimensiones pueden ser obtenidos a través de tres abordajes: “data-driven” (o guiados por los datos), precios de mercado o métodos normativos (Sehnbruch, 2020). En este estudio proponemos una combinación de abordajes normativos (a partir de la selección de las variables) y guiados por los datos (“data driven”) (a partir de la utilización de técnicas estadísticas de agregación de las diferentes dimensiones y variables). En ese sentido, gran parte del debate está en la selección de la técnica que permitan agregar de la manera más robusta posible los indicadores seleccionados. Aquí proponemos la utilización del Análisis de Componentes Principales Policórico (ACPP) para la agregación de las variables. Esta técnica considera las correlaciones policóricas y tetatróricas en la construcción del índice (y al basarse en variables ordinales permite introducir el conocimiento normativo previo en la construcción de las categorías de cada variable incluida). Posteriormente, se analiza el comportamiento del índice tanto por cada componente por separado como en el agregado global mediante regresión por mínimos cuadrados. Los principales resultados obtenidos sugieren: a) que el status migratorio y el género afectan la calidad del empleo, aportando evidencia empírica en dirección de que estos factores son relevantes en la producción y reproducción de las desigualdades; b) escalas subnacionales del país presentan diferentes calidades del empleo (Noroeste peor calidad y Buenos Aires mejor) aportando evidencia empírica sobre la heterogeneidad de la estructura productiva de Argentina.

Palabras Clave: Calidad del empleo, Índice Multidimensional, Demografía

¹e-mail: ebologna@gmail.com

²e-mail: pablogomez@unc.edu.ar

³e-mail: miglesias@unc.edu.ar

Expecting the unexpected: economic growth under stress

Gloria González-Rivera ¹, Vladimir Rodríguez-Caballero ², Esther Ruiz ³¹

¹ Department of Economics, University of California, Riverside,

² Department of Statistics, ITAM (Mexico) and CREATES, Aarhus University (Denmark),

³ Department of Statistics, Universidad Carlos III de Madrid (Spain)

Large and unexpected moves in the factors underlying economic growth should be the main concern of policy makers aiming to strengthen the resilience of the economies. We propose measuring the effects of these extreme moves in the quantiles of the distribution of growth under stressed factors (GiS) and compare them with the popular Growth at Risk (GaR). In this comparison, we consider local and global macroeconomic and financial factors affecting US growth. We show that GaR under estimates the extreme and unexpected fall in growth produced by the COVID19 pandemic while GiS is much more accurate. Figure 1 displays differences between GaR and GiS.

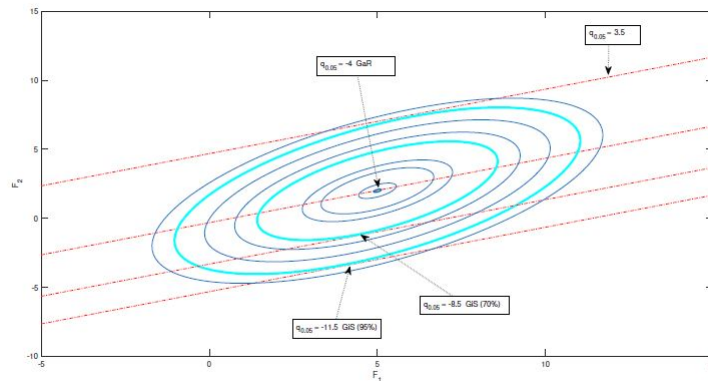


Figura 1: GaR and GiS. The red dash-dot lines are the iso- 5% quantile growth lines for different values of the 5% quantile. There are four iso-quantile lines that correspond to $q_{0,05}(y_{t+1} | F_t) = 3.5, -4, -8.5$ and -11.5 . The blue ellipses are the contours of the bivariate probability density function of the factors. The highlighted blue ellipses are the 70% and 95% probability contours. The GaR is the value of the iso- 5% quantile line at the center of the ellipse, which are the averages of the factors. The GiS is the value of the iso- 5% quantile line that is tangent to the 70% probability contour (GiS = -8.5) or 95% probability contour (GiS = -11.5).

Consult the working paper [here](#).

Some relevant references:

1. Adrian, T., Boyarchenko, N., Giannone, D., 2019a. Vulnerable growth. American Economic Review 109, 1236-1289.
2. Ando, T., Tsay, R., 2011. Quantile regression models with factor-augmented predictors and information criterion. Econometrics Journal 14, 1-24.
3. González-Rivera, G., Maldonado, J., Ruiz, E., 2019. Growth in stress. International Journal of Forecasting 35, 948-966.
4. Rodríguez-Caballero, C., Caporin, M., 2019. A multilevel factor approach for the analysis of CDS commonality and risk contribution. Journal of International Financial Markets, Institutions & Money 63, 101-144.

Palabras Clave: Growth vulnerability, Multi-level factor model, Stressed growth.

¹e-mail: gloria.gonzalez@ucr.edu, vladimir.rodriguez@itam.mx, ortega@est-econ.uc3m.es

Fenómenos dinámicos: paradigmas para su representación

Mirta L. González, Alberto H. Landro¹

Centro de Investigaciones en Econometría, IADCOM, UBA

Aceptar la concepción clásica -determinística- implica asumir como punto de partida una axiomática según la cual todo sistema dinámico es formalizable en el supuesto que, en el límite, su comportamiento obedece a una trayectoria que es la consecuencia necesaria de la influencia de infinitos factores que constituyen su estructura causal y cuya modelización matemática, inevitablemente incompleta, genera un tipo de incertidumbre equiparable a una aleatoriedad sinónimo de ignorancia. Por el contrario, la concepción termodinámica -aleatorista- implica sustituir el axioma de existencia de una trayectoria por el supuesto de comportamiento aleatorio que puede generar eventualmente ciertas regularidades locales a nivel macroscópico y, en consecuencia, sustituir la interpretación clásica de azar ignorancia por la de azar absoluto. El objetivo de este trabajo es contribuir a la solución de la histórica cuestión de los fundamentos de la naturaleza de la aleatoriedad inherente al comportamiento de los fenómenos dinámicos, a partir de la demostración formal de la verdad o falsedad de estas proposiciones utilizando los argumentos que proporciona la teoría de la complejidad algorítmica y de la característica de infinitamente numerable de la estructura causal de un sistema económico dinámico y la posibilidad de la asimilación de su modelización al ámbito formalista de la matemática. A partir de los argumentos de las tesis de Church-Turing-Chaitin, se concluyó en la indemostrabilidad tanto de la tesis aleatorista como de la de su contraria, la tesis determinística y, en consecuencia, en la demostración de la imposibilidad de resolver formalmente la disyuntiva histórica determinismo-aleatorismo.

Palabras Clave: Fenómenos dinámicos, aleatorismo, determinismo, tesis de Church-Turing, complejidad algorítmica, Chaitin

¹e-mail: alandroar@yahoo.com.ar; mgonza@economicas.uba.ar

Estimación del componente cíclico del PIB de Uruguay mediante filtros pasabanda y HP

Silvia Rodríguez-Collazo, Federico Molina.¹

*Instituto de Estadística (IESTA), Facultad de Ciencias Económicas y Administración
Departamento de Métodos Cuantitativos, Universidad de la República*

A partir de la Ley de Urgente Consideración aprobada en 2020 en el que se estipula que en la ejecución presupuestal, se presentará el déficit fiscal ajustado por el ciclo, la preocupación y debate sobre las alternativas de estimación de la brecha de producto medida a través del componente cíclico del PIB trimestral de Uruguay ha cobrado nueva vigencia e interés.

La metodología que sustenta los filtros aplicados y este trabajo se apoya en la hipótesis de que este componente cíclico estimado representa adecuadamente al ciclo de referencia. Recoge y caracteriza a las fluctuaciones identificadas por Burns y Mitchell (1946) como fluctuaciones cíclicas.

En este trabajo se presentan un conjunto de estimaciones del componente cíclico del PIB de Uruguay para la serie observada que se inicia el primer trimestre de 1997 hasta el primer trimestre de 2021, adicionalmente se consideran predicciones que pueden alcanzar hasta el cuarto trimestre de 2024. Las estimaciones se obtendrán a partir de la aplicación de tres filtros, el propuesto por Hodrick- Prescott (1981), el filtro pasa banda de Baxter - King (1999) y finalmente el filtro pasa banda diseñado por Christiano-Fitzgerald (2003).

El trabajo se traza los siguientes objetivos: explicitar las propiedades de los ciclos estimados a partir de los filtros utilizados, así como la capacidad de captar adecuadamente la señal cíclica, en términos de su distancia con un filtro ideal. Comparar las estimaciones obtenidas al aplicar los diversos filtros, caracterizar y analizar las señales obtenidas en términos de signo y magnitud de la distancia al componente tendencial. Finalmente mostrar la sensibilidad de los resultados obtenidos a partir de proyecciones alternativas sobre la trayectoria del PIB para el período 2021-2022, 2021-2024 provenientes de diversas fuentes, entre ellas las proyecciones del Ministerio de Economía y Finanzas (MEF) incluidas en la última Rendición de Cuentas.

Se repasan trabajos antecedentes para Uruguay, Rodríguez y Badagián (2004), Rodríguez, et al (2007a,b,c), Álvarez y da Silva (2008) en lo que refiere a la aplicación de estos filtros para estimar el ciclo de referencia del PIB en el pasado. En esta ocasión el énfasis es en la actualización de los resultados para la serie de IVF del PIB Base 2016 =100 así como los ciclos resultantes al introducir proyecciones alternativas sobre la evolución del producto.

En este trabajo se aplican un conjunto acotado de metodologías de estimación del ciclo de referencia, no obstante la variabilidad en los resultados obtenidos permite alertar sobre la importancia de la elección de la metodología para la estimación de la señal cíclica. El papel relevante que tiene y tendrá la transparencia a la hora de explicitar la metodología, pero el nivel de transparencia requerido es aquel que permita la reproducibilidad de los resultados. Con ello será posible hacer efectivo un seguimiento real al resultado fiscal estructural presentado por el MEF en instancias de las sucesivas rendiciones de cuentas y balances de ejecución presupuestal, así como la comparación entre las metas planteadas y los resultados alcanzados.

Palabras Clave: Brecha de producto, ciclo, filtro Hodrick-Prescott, filtro Baxter-King, filtro Christiano-Fitzgerald, filtros pasabanda

¹e-mail: silvia.rodriguez@fcea.edu.uy, federico.molina@fcea.edu.uy

Una propuesta dinámica del Índice Riesgo País para Uruguay

Andrés Sosa ¹

Instituto de Estadística, Departamento de Métodos Cuantitativos, Universidad de la República

En el trabajo se introduce un modelo estocástico que nos permite obtener un índice de riesgo país dinámico para Uruguay. La propuesta se basa en el enfoque de intensidad del riesgo de crédito, es decir, el riesgo de incumplimiento se modela como el primer salto de un proceso con intensidad estocástica. Se comparan dos propuestas de modelos para estimar el diferencial de rendimiento entre los bonos soberanos de Estados Unidos y los de Uruguay. La dinámica de las tasas de interés se modela mediante un modelo afín multidimensional y para calibrarlo se utiliza el filtro de Kalman mediante el algoritmo de esperanza-maximización (EM). Las tasas de interés de Estados Unidos constituyen una parte de la entrada del modelo, así como los precios de los principales bonos nacionales.

Trabajo en conjunto con Ernesto Mordecki.

Palabras Clave: Índice Riesgo País, deuda soberana, Filtro de Kalman, Algoritmo Expectation Maximization.

¹e-mail: andres.sosa@fcea.edu.uy

Tema 8: Educación

Análisis discriminante para el modelado del rendimiento matemático en estudiantes de Ingeniería

Humberto Closas, Edgardo Arriola, Mariela Amarilla, Carina Jovanovich¹

Facultad Regional Resistencia, Departamento de Materias Básicas, Universidad Tecnológica Nacional

La determinación y el análisis de los aspectos que intervienen en el rendimiento de los estudiantes es una problemática compleja de abordar, en razón del número de variables que participan y de las interacciones que entre ellas pueden presentarse.

En la región nordeste de Argentina, como en otras zonas de este país y de América Latina, debido principalmente a la escasa preparación que poseen los alumnos al llegar a la Universidad, sumado a la fragilidad que se observa en su orientación vocacional, poco tiempo después de ingresar, muchos de ellos abandonan los estudios o deciden cambiar de carrera.

En atención a lo que antecede, el objetivo principal de este estudio consiste en elaborar mediante la técnica multivariada denominada *análisis discriminante*, un modelo que explique de qué manera se relacionan ciertas variables personales (autoconcepto académico y estrategias de aprendizaje) y contextuales (aspectos sociofamiliares y elementos del clima de clase) con el rendimiento académico de estudiantes universitarios, en el ámbito de una asignatura del área de Matemática.

La variable dependiente del modelo fueron las calificaciones obtenidas por los estudiantes en el examen parcial de Análisis Matemático I realizado en junio de 2019. Si bien la evaluación de los conocimientos en un procedimiento verdaderamente complejo, hemos utilizado las mencionadas calificaciones por dos motivos: a) las ventajas que derivan del uso de criterios estandarizados (p. ej., fiabilidad, estabilidad y comparabilidad), y b) se consideraron un ratio representativo del aprendizaje producido por la interacción con los contenidos desarrollados, por cierto, hasta el momento de la instancia examinadora.

La muestra estuvo compuesta por 142 jóvenes, con una media de 19.75 años ($DE = 1.42$), que en el año antes mencionado se encontraban matriculados en el primer curso de las carreras: Ingeniería en Sistemas de Información, Ingeniería Electromecánica e Ingeniería Química, que se desarrollan en la sede central de la Facultad Regional Resistencia de la Universidad Tecnológica Nacional (Argentina).

El diseño metodológico es inicialmente de naturaleza *no experimental* y, en segunda instancia, de tipo *explicativo*, debido al objetivo propuesto. Por la clase de información que se provee, el modo de reunir la y el marco donde se aplicaron las pruebas, el estudio es de estilo *descriptivo mediante encuesta en trabajo de campo*. En razón de cómo se miden y analizan los datos, es una investigación de línea *cuantitativa*. Debido a la instancia de recolección de la información, el trabajo es de corte *transversal*. Dado el interés por analizar las asociaciones entre las variables que participan, así como de proyectar recomendaciones que resulten viables y sustentables en el tiempo, el diseño es de perfil *correlacional y prospectivo*.

En la fase empírica de la investigación, además de los análisis estadísticos que posibilitaron evaluar la confiabilidad de los instrumentos utilizados para medir las variables independientes, fue posible contrastar que el modelo discriminante estimado se ajusta a los datos muestrales y resultaría de utilidad para clasificar nuevos individuos en los grupos de pertenencia definidos a priori, siempre que se disponga de los valores en las variables independientes (finalidad predictiva de la ecuación que se propone), y que el proceso de enseñanza sea semejante al que se desarrollaba en el momento de aplicar las pruebas y recoger las observaciones.

Finalmente, se plantean estrategias pedagógicas de mediación, tanto preventivas como correctivas, a partir de las causas (variables de tipo motivacional, cognitiva, sociológica y educativa) que, según los resultados de esta investigación, afectan el rendimiento en modo presencial de la población de estudiantes de la que deriva la muestra seleccionada.

Palabras Clave: Modelización estadística, Rendimiento matemático, Factores personales y contextuales, Estudiantes universitarios.

¹e-mail: hclosas@hotmail.com, earriola2006@yahoo.com.ar, profe.amarilla@gmail.com, carinajovanovich@gmail.com

Modelo predictivo de las relaciones entre variables personales, educativas y el rendimiento en Contabilidad

H. Closas, I. de Castro, N. Franchini, R. Cruz, A. Dusicka, L. Kuc¹

Facultad de Ciencias Económicas, Departamento de Matemática – Departamento de Contabilidad, Universidad Nacional del Nordeste

Desde hace tiempo, el rendimiento académico en las aulas universitarias es altamente preocupante debido, entre otras razones, a la notable distancia que existe entre los conocimientos reales que poseen los estudiantes y los que resultan necesarios para lograr un correcto desempeño académico. A su vez, por tratarse de una problemática multidimensional, es posible abordarla desde distintos enfoques, también teniendo en cuenta diferentes razones, así como sus vinculaciones e implicancias.

En este marco, el objetivo del presente trabajo radica en elaborar un modelo, a través de la técnica *análisis discriminante*, que permita tanto explicar como predecir, de qué manera ciertas variables intrínsecas y extrínsecas se encuentran relacionadas con los resultados en una asignatura del área contable.

La muestra estuvo conformada por 110 jóvenes, con una media de 19.84 años ($DE = 1.78$), que ingresaron, en el curso académico previo a la declaración, por parte de la OMS, de la pandemia de COVID-19, a la Facultad de Ciencias Económicas de la Universidad Nacional del Nordeste (Argentina).

El estudio responde, en virtud del objetivo que persigue, a un diseño explicativo y predictivo. Es también una investigación de corte transversal, de estilo descriptivo mediante encuesta en trabajo de campo y de línea cuantitativa. En razón del interés por analizar las asociaciones entre las variables, así como de realizar recomendaciones que resulten útiles para mejorar el rendimiento de nuevos jóvenes ingresantes, la metodología posee un claro perfil correlacional y prospectivo.

Se utilizaron variables motivacionales, cognitivas y de contexto para modelar el fenómeno del rendimiento, cuyos indicadores fueron las calificaciones parciales promedio obtenidas por los estudiantes que integraron la muestra productora de datos. La decisión de incluir o excluir determinadas variables en el estudio fue un hecho que inevitablemente se presentó condicionado por diversas circunstancias, tales como, la característica multifactorial del constructo, el consenso científico internacional, el criterio subjetivo de los investigadores y la viabilidad del estudio en un tiempo razonable de desarrollo.

Respecto de las calificaciones (variable dependiente en el modelo), las mismas fueron seleccionadas en razón de que son el criterio legal, formal y social del rendimiento en el ámbito de los centros educativos, además de ser uno de los indicadores más utilizado en las investigaciones sobre esta temática, aunque se reconoce sus limitaciones objetivas y subjetivas.

En la etapa empírica, los análisis estadísticos implementados inicialmente permitieron, por un lado, conocer el grado de fiabilidad de los instrumentos utilizados para medir las variables independientes y, por otro, confirmar que las mismas se encuentran normalmente distribuidas, uno de los supuestos previos de la técnica aplicada.

Posteriormente, fue posible contrastar que el modelo discriminante estimado se ajusta a los datos muestrales y resultaría de utilidad para clasificar nuevos individuos en los grupos de pertenencia definidos a priori, a partir de los valores en las variables independientes y, por cierto, en un contexto académico similar al que existía al momento de aplicarse los instrumentos de medición.

Por último, teniendo en cuenta la función propuesta, se plantearon medidas de intervención educativa, relacionadas con el *autoconcepto académico*, las *estrategias de aprendizaje* y el *proceso de enseñanza*, con el objeto de aportar posibles soluciones al problema del rendimiento en modo presencial en la asignatura objeto de interés.

Palabras Clave: análisis discriminante, rendimiento en Contabilidad, variables motivacionales, cognitivas y de enseñanza, estudiantes universitarios.

¹e-mail: hclosas@hotmail.com, idecastro@eco.unne.edu.ar, noeliabfranchini@gmail.com, rcruz@eco.unne.edu.ar, mad2607@yahoo.com, lucianakuc@gmail.com

Título Implementación de metodologías de análisis multivariado para caracterizar estudiantes en el aprendizaje de fracciones matemáticas

Marcela L. Natal(1), María Laura Patat (1), Mauro A. E. Chaparro(1,2), María M. Richard's (3) ¹

*1*Centro Marplatense de Investigaciones Matemáticas- Universidad Nacional de Mar del Plata (CEMIM-UNMDP) *2* Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) *3* Instituto de Psicología Básica, Aplicada y Tecnología (IPSIBAT – CONICET – UNMDP) , Departamento, Universidad

Este trabajo surge de un Proyecto Interfacultades (Facultad de Psicología- Facultad de Ciencias Exactas y Naturales) en la UNMdP. Dicho proyecto analiza el aprendizaje de las fracciones matemáticas, su relación con procesos de control cognitivo e inteligencia fluida en estudiantes que finalizan la escolaridad primaria. La competencia con las fracciones es importante para aprender más matemática avanzada. Se obtuvo una muestra incidental no probabilística de 97 estudiantes provenientes de tres escuelas primarias públicas y una privada de la ciudad de Mar del Plata. La información obtenida a partir de la implementación de distintas pruebas referidas a conocimientos de matemática previos e inteligencia fluida, generó las 11 variables del conjunto de datos analizado. El objetivo de este trabajo es comparar y aplicar diferentes metodologías estadísticas para clasificar los 97 estudiantes que representan los procesos cognitivos, sus competencias matemáticas previas y el conocimiento de fracciones. Para ello, se implementaron metodologías multivariadas tales como Análisis de componentes principales (PCA), análisis discriminante, análisis de clúster jerárquico y difuso. El clúster difuso tiene por finalidad detectar aquellos individuos que presentan características de transición entre los grupos obtenidos por el clúster jerárquico. Se compararon las clasificaciones obtenidas por las diferentes metodologías. Se obtuvieron cuatro clústers de estudiantes con características identificables e interpretables desde la psicología. Con este trabajo se espera incentivar el empleo de métodos estadísticos multivariados en el área de la psicología que permitan establecer tipologías o perfiles de los estudiantes en el aprendizaje de la matemática y que, además éstos se utilicen para la toma de decisiones a nivel de políticas públicas en educación.

Palabras Clave: Clúster jerárquico, clúster difuso (fuzzy c-means clustering), PCA, análisis discriminante.

¹e-mail: mnatal@mdp.edu.ar, mlpatat@mdp.edu.ar

Prácticas asociadas al aprendizaje de las integrales curvilíneas en las Ciencias de la Salud

Christiane Ponteville, Myriam Nuñez, María José Castro¹

Cátedra de Matemática, Departamento de Fisicomatemática, Facultad de Farmacia y Bioquímica, UBA

El objetivo de este trabajo es identificar prácticas asociadas en los problemas que involucran formas diferenciales, integrales curvilíneas e integrales dobles. Para ello se identificaron los errores y se detectaron los obstáculos en la resolución del examen correspondiente a la asignatura Matemática de las carreras de Farmacia y Bioquímica de la Facultad de Farmacia y Bioquímica de la Universidad de Buenos Aires durante el primer y segundo ciclo lectivo del año 2020.

La importancia de este análisis se basa en el lugar fundacional que poseen los conceptos matemáticos con los que se formula y se da sustento a las teorías relacionadas con el área de las ciencias de la salud.

Los cursos se dictaron en forma virtual, con clases sincrónicas y asincrónicas. Los exámenes se rindieron en la virtualidad a través del campus virtual de la asignatura. El diagnóstico de los obstáculos nos permitirá realizar un seguimiento periódico de los aprendizajes de concepciones matemáticas fundamentales vinculadas al cálculo diferencial, integral y vectorial.

Los ejercicios involucrados son del tipo: **T1**- resolución de una integral curvilínea de una forma diferencial exacta sobre un camino cerrado; **T2**- resolución de una integral curvilínea de una forma diferencial exacta; **T3**- verificación de las hipótesis del Teorema de Green y resolución de la integral curvilínea a través de una integral doble; **T4**- resolución de una integral curvilínea de una forma diferencial sobre un camino cerrado; **T5**- resolución de una integral curvilínea de una forma diferencial.

De las 277 resoluciones analizadas 30 corresponden al tipo **T1**(10, 8%), 64 al tipo **T2**(23, 1%), 82 al tipo **T3**(29, 6%), 66 al tipo **T4**(23, 8%) y 35 al tipo **T5**(12, 7%).

Algunos de los errores que detectamos en las resoluciones de los ejercicios fueron:

- Objetos matemáticos involucrados: curvas en el plano, forma diferencial, integral curvilínea.
- Conceptos previos: operatoria algebraica, cálculo de derivadas, integrales definidas y funciones trigonométricas.
- Interpretación de las hipótesis y tesis de un teorema; vinculación de las representaciones de una integral curvilínea y una doble.
- Comprensión de la función potencial: condición necesaria para su existencia, asociación con una forma diferencial, vinculación con la integral.

Para el análisis de los datos se utilizaron la prueba de Chi-Cuadrado y el Análisis Factorial de Correspondencias.

Algunas observaciones que obtuvimos al analizar los diferentes tipos de ejercicio fueron las siguientes: diferentes maneras de validar el resultado (los errores se presentaron sólo en la justificación siendo el más común el relacionado al cálculo del diferencial al aplicar el método de sustitución); diversos errores procedimentales en el cálculo de la función potencial y falta de interpretación de los límites como puntos en el plano; falta de la verificación del cumplimiento de las hipótesis del teorema de Green (algunos verifican la tesis cuando no era lo pedido); identificación incorrecta de los límites de integración; la representación gráfica de la región correspondiente no fue realizada por la mayoría de los estudiantes.

Concluimos que debemos ser más explícitos al momento de redactar los enunciados de los ejercicios para poder evaluar en forma integral los aprendizajes de los estudiantes. Además, debemos considerar la realización de un trabajo áulico que relacione los lenguajes matemático, gráfico y simbólico.

Palabras Clave: prácticas asociadas, integrales curvilíneas, Teorema de Green, obstáculos

¹e-mail: chponteville@gmail.com, myriam@ffyba.uba.ar, mjcastrogonzalez@gmail.com

Aplicación de Minería de Textos a la evaluación estudiantil universitaria: desafíos y posibilidades

Daniel Alessandrini¹

Unidad de Enseñanza, Facultad de Ingeniería, Universidad de la República

La evolución de las tecnologías de la información y comunicación a nivel global ha permitido extraer información de muy diversas fuentes, con volúmenes de datos cada vez mayores. Particularmente, gracias a la minería de textos hoy es posible obtener una gran variedad de información de interés para distintas disciplinas del saber, sacándole provecho desde gobiernos y entidades públicas hasta organizaciones, empresas y la academia. En el ámbito de la enseñanza universitaria, las evaluaciones al cuerpo docente resultan de interés no solo para poder entender donde hay problemas y como mejorarlos, sino también para quiénes evalúan y diseñan las currículas de las diferentes carreras, e incluso para quienes se encargan de la gestión universitaria.

En la actualidad, la Facultad de Ingeniería de la Universidad de la República (FIng) utiliza la información proveniente del formulario F5 del Sistema de Evaluación Docente en la función Enseñanza (SEDE) para efectuar devoluciones a los diferentes Institutos, específicamente con las preguntas cerradas, es decir aquellas que presentan opciones predeterminadas de respuesta. Hasta el momento las preguntas abiertas que existen en los formularios no han sido utilizadas. Sin embargo, diferentes trabajos muestran la importancia de incorporar análisis específicos a este tipo de respuestas para ordenar prioridades en el dictado de cursos, qué cambiar de los mismos, además de darle a FIng herramientas de evaluación más precisas [1]. A pesar de ser corpus textuales relativamente pequeños -comparado con otras aplicaciones, p.ej. análisis de redes sociales, de documentos de gran porte, etc.- éstos presentan complejidades propias: muchas veces, en una misma sentencia los estudiantes dejan por escrito información diferente (p.ej. sugerencias de utilidad o comentarios negativos, incluso en una misma oración) [2]. Adicionalmente, se ha encontrado en estudios preliminares una baja coincidencia con diccionarios de uso general, haciendo que sea difícil llevar adelante técnicas como clasificar o asignar polaridad a palabras y comentarios.

De este modo, se presentarán resultados preliminares respecto a la creación de un diccionario específico para este tipo de corpus, considerando tanto diccionarios existentes como también uno específicamente creado como complemento de los anteriores para mejorar polaridad por *token*, minimizando positivos o negativos que en contextos diferentes tienen ese sentido (y en estas encuestas no) [3]. Se espera a futuro generalizar los resultados encontrados, como así también la incorporación de otras técnicas textuales (modelización de tópicos, etc.) para enriquecer el análisis.

Referencias

- [1] Swapna Gottipati, Venky Shankararaman, and Jeff Rongsheng Lin. Text analytics approach to extract course improvement suggestions from students' feedback. *Research and Practice in Technology Enhanced Learning*, 13(1):6, December 2018.
- [2] Gokarn Ila Nitin, Gottipati Swapna, and Venky Shankararaman. Analyzing educational comments for topics and sentiments: A text analytics approach. In *2015 IEEE Frontiers in Education Conference (FIE)*, pages 1–9, Camino Real El Paso, El Paso, TX, USA, October 2015. IEEE.
- [3] Francisco José Martínez. Análisis de Sentimiento en Twitter de las principales Compañías del Sector Asegurador Español. Master's thesis, Valencia, España, 2017.

Palabras Clave: minería de textos, análisis multivariado, encuestas estudiantiles

¹e-mail: dalessandrini@fing.edu.uy

Segundo año de pandemia. Sigue presente el Burnout académico en estudiantes universitarios?

Adriana D'Amelio¹, Lucía Nardechia², Johana Tari³

^{1,2} *Universidad Nacional de Cuyo*

³*Dirección General de Escuela. Mendoza*

adriana.damelio@fce.uncu.edu.ar

RESUMEN

En los tiempos actuales transitando el mundo por la pandemia Covid-19 los estudiantes en las universidades han ido adaptándose al cambio de la presencialidad a la virtualidad. Aunque las estrategias y las características de la continuidad pedagógica, al principio de la pandemia mostró variaciones de acuerdo a la disponibilidad de recursos de las Universidades, en la actualidad muchas de esas dificultades se fueron subsanando.

Lo anterior puede contribuir en los estudiantes, sensaciones de no poder dar más de sí mismo, tanto física como psíquicamente, desvalorización, pérdida del interés del valor frente al estudio y dudas crecientes acerca de la propia capacidad para realizarlo. La presencia simultánea de estas manifestaciones es conocida como "síndrome de burnout académico".

Finalizado el primer semestre del 2020 se hizo una investigación con estudiantes de la carrera de Contador de la facultad de Ciencias Económicas de la Universidad Nacional de Cuyo a los que se les aplicó un cuestionario, Maslach Burnout Inventory (MBI) que se utiliza para medir dicho síndrome. Se realizaron 250 encuestas que mostraron la presencia de Burnout en un 20% en estudiantes de los primeros años y en un 40% con puntajes cercanos. Es por ello que para continuar la investigación se decidió tomar de nuevo la encuesta a estudiantes de 2º año de la carrera de Contador en 2021.

Con esta investigación se pretende detectar el comportamiento de los estudiantes en este segundo año de pandemia, nuestra hipótesis es que algunos ya se han adaptado y encontraron la manera de estudiar en estas circunstancias, otros no y su nivel de Burnout creció y evaluar los resultados desde metodologías estadísticas adecuadas.

Palabras claves: Síndrome de *burnout*, pandemia, estudiantes, universidad.

Tema 9: Enseñanza

Experiencia de utilización de Google Street View como herramienta de muestreo digital para la enseñanza de estadística.

Nicolás E. Tamburi¹ y M. Cecilia Tiecher²

1. INBIOSUR / Departamento de Matemática, Universidad Nacional del Sur.

2. SAPRHA

En el marco de la pandemia covid-19 se propuso una actividad para estimular el interés y compromiso por parte del alumnado en la materia Bioestadística (dictado 2021, a cargo del Profesor Ricardo Camina). En la ciudad de Bahía Blanca, existe una gran abundancia de palomas domesticas (*Columba livia*) que generan problemas sanitarios y preocupación por potenciales problemas zoonóticos. Aprovechando la herramienta gratuita Google Street View (GSV) que ofrece una colección de imágenes a lo largo del tiempo se propuso a los alumnos comparar la abundancia de esta especie en las fachadas de las edificaciones del centro de la ciudad en 2015 y 2020. Se realizó un video explicativo que los condujo a realizar un muestreo apareado en 20 sitios de la ciudad contabilizando la abundancia de palomas. Considerando sus competencias científicas se les mencionó que miembros de la cátedra y del INBIOSUR estamos desarrollando una investigación similar para comprender la dinámica poblacional de esta especie invasora.

El análisis de los datos fue llevado a cabo por el docente realizando una demostración interactiva sincrónica. La actividad sirvió como disparador para discutir varios temas curriculares entre ellos, como generar hipótesis apropiadamente, como diseñar un muestreo, diferencias entre muestreo apareado e independiente, nivel de significación o error, *p value*, transformaciones, etc. Tal como era esperable según datos previos se observó un aumento significativo en el número de palomas en 2020 respecto de 2015.

Los alumnos fueron evaluados con un examen utilizando la plataforma moodle-uns y se realizó una devolución constructiva del mismo a la semana siguiente. Finalmente completaron una encuesta anónima y optativa que mostró la posible utilidad de este muestreo digital como herramienta didáctica. La cantidad de alumnos se triplicó durante la explicación del TP y aumento 2.7 veces durante la devolución constructiva del examen respecto del promedio de las clases precedentes. La calificación media de las alumnas y alumnos fue de 7.7; solo una alumna no alcanzo los objetivos y debió recuperar. De los 16 alumnos que respondieron la encuesta el 93.75 % manifestaron que no tuvieron dificultades para realizar la actividad, el 100 % reconoce que les sirvió en diferente grado para comprender mejor los conceptos relacionados con las pruebas de hipótesis. Consideraron la complejidad de la evaluación como media y un 87 % recomendó utilizar esta actividad para futuros dictados de esta materia.

Las imágenes de GSV ya han sido utilizadas para estudiar la distribución y abundancia de varias especies vegetales y animales, así como para estudios urbanísticos. Sumado a las imágenes satelitales de distribución gratuita que se dispone actualmente son un recurso aprovechable también para la enseñanza de estadística básica. A nuestro entender, la toma de datos por parte del alumno es sumamente importante en tanto genera un sentido de pertenencia de la pregunta que luego lleva a un uso más apropiado de las técnicas estadísticas utilizadas para el estudio.

Palabras Clave: Enseñanza de la estadística, Muestreo, Prueba de hipótesis

¹e-mail: ntamburi@uns.edu.ar

²

Identificación de prácticas frecuentes en la resolución de problemas de cálculo de probabilidades

Daniela Parada, Christiane Ponteville, Myriam Núñez¹

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

Este trabajo tiene su origen en la identificación y caracterización de los tipos de prácticas utilizadas por los alumnos en el aprendizaje de la probabilidad conjunta y condicional, de la independencia estadística y, transversalmente a estos, las dificultades que encuentran en la elección y utilización de modos de representación adecuados. Para ello, se analizaron las resoluciones de los ejercicios planteados en el examen parcial.

La relevancia del trabajo se apoya en la convicción de que dichos conceptos son fundamentales para la formación estadística de futuros profesionales de las Ciencias de la Salud, especialmente, en la etapa de inferencia como en las pruebas de hipótesis. Por ello, resulta de interés identificar las prácticas frecuentes que los estudiantes utilizan en relación con estos conceptos, para poder delinear estrategias de enseñanza y evaluación que los puedan hacer conscientes de las dificultades habituales y para que puedan afrontar los problemas probabilísticos con herramientas adecuadas.

La asignatura Bioestadística de las carreras de Farmacia y Bioquímica de la Facultad de Farmacia y Bioquímica de la Universidad de Buenos Aires se cursa entre el cuarto y el sexto cuatrimestre. Los alumnos acceden a cursarla luego de haber aprobado las asignaturas de Matemática en las que se presentan contenidos vinculados a elementos del cálculo diferencial e integral en una y varias variables.

La guía de trabajos prácticos de la materia está organizada siguiendo los lineamientos habituales de la enseñanza de la Probabilidad y la Estadística en cursos de grado. La introducción a los fundamentos del cálculo de probabilidades es relativamente rápida (una unidad de las ocho disponibles) y se alcanzan a presentar nociones de fenómenos aleatorios, espacio de resultados, sucesos, probabilidad, frecuencia relativa, cálculo de probabilidades (entre las que se destaca la condicional y la intersección) e independencia.

En los exámenes, los ejercicios de esta unidad generalmente proponen el cálculo de probabilidades (usualmente, condicionales, intersecciones, totales) a partir de enunciados coloquiales que exigen una tarea de interpretación y, a veces, de la elección de un registro de representación adecuado. En ocasiones, los sucesos se denotan y definen explícitamente en el enunciado, aunque no es lo más frecuente.

Del análisis de las resoluciones de los estudiantes, se identifican algunas de las prácticas más frecuentes de acuerdo con su aparición sistemática. A saber: los sucesos son definidos como categorías en lugar de ocurrencias lo que, en ocasiones, obstruye el cálculo de las probabilidades posteriores; se confunde la probabilidad condicional con la de la intersección, especialmente cuando el enunciado está dado coloquialmente y en ausencia de identificadores como “dado que” o “sabiendo que”; se identifica la probabilidad simple con la condicional, especialmente cuando hay una definición tácita de los sucesos; se confunde la probabilidad condicional con su inversa, esto es, se invierte el rol del suceso condicionado y el condicionante; se asume la independencia, aun en contextos en donde es requerido probarla.

Además, se observa que no es frecuente que los estudiantes utilicen representaciones adecuadas como tablas de contingencia, diagramas de árbol o diagramas de Venn para la resolución de problemas, a la vez que se observa que, en ocasiones, sus estrategias se ven fuertemente influenciadas por la forma en la que el enunciado es dado.

Con base en el análisis realizado, observamos que es necesario explicitar con mayor claridad los enunciados de los ejercicios propuestos en el examen como así también hacer hincapié en su presentación y desarrollo durante las clases prácticas.

Palabras Clave: enseñanza de la probabilidad, identificación de prácticas, probabilidad condicional, probabilidad conjunta

¹e-mail: dparada@cbc.uba.ar, chponteville@gmail.com, myriam@ffyb.uba.ar

ESTRATEGIAS PARA LA ENSEÑANZA DE LA ESTADÍSTICA EN CIENCIAS MEDIANTE EL SOFTWARE LIBRE R

Claudia Buogiorno ^{1 2}, Claudia Castro-Kuriss ¹, Osmar Vera ^{2, 1}

¹ *Departamento de Matemática, Instituto Tecnológico de Buenos Aires (ITBA) , Ciudad Autónoma de Buenos Aires, Argentina.*

² *Departamento de Ciencia y Tecnología, Área Matemática, Universidad Nacional de Quilmes (UNQ), Pcia. de Buenos Aires, Argentina.*

La Estadística es considerada actualmente una ciencia en sí misma que conlleva una forma de abstracción diferente, por ello para su enseñanza es necesario establecer estrategias que permitan su comprensión y su aplicación en problemas prácticos. El empleo de un software libre como el R potencia las posibilidades de aplicación. Es necesario tratar nuevos temas en el aula de clase (simulaciones, estimaciones bootstrap, gráficos novedosos) y omitir temas obsoletos (tablas, interpolaciones, aproximaciones innecesarias). La computación está cambiando la forma de hacer estadística y la forma de enseñarla junto con la virtualidad que se ha instalado actualmente como método de dictado. Esto necesariamente lleva a lo que se conoce hoy como Ciencia de los Datos. Los autores de este trabajo proponen replantear la enseñanza actual mediante el uso de las estrategias didácticas del Proyecto Mosaic para un aprendizaje constructivista de los conceptos estadísticos mediante el software libre R. Dicho proyecto, fundado por una comunidad de educadores, se focaliza en aspectos del trabajo cuantitativo que los estudiantes de grado de carreras de ciencia, tecnología e ingeniería, necesitarán en sus vidas profesionales y que usualmente hoy reciben con una formación aislada a esa realidad profesional. El Proyecto Mosaic ha creado la correspondiente librería en R para simplificar el uso de este software en la etapa introductoria e incluye también interesantes conjuntos de datos reales. El objetivo de este trabajo es mostrar las técnicas gráficas incluidas en la librería Mosaic aplicadas a conjuntos de datos relevantes para los estudiantes de las carreras mencionadas junto con las técnicas estadísticas necesarias para investigarlas e interpretarlas, que los autores ya han llevado a la práctica con éxito en varios cursos de su especialidad.

Palabras Clave: Enseñanza de la estadística, Software R, Mosaic Project.

¹ e-mail: cbuongiorno@itba.edu.ar, ccastro@itba.edu.ar, overa17@gmail.com

IDENTIFICACIÓN DE LAS ARGUMENTACIONES ASOCIADAS AL MODELO DE REGRESIÓN LINEAL SIMPLE

NUÑEZ, M, BONADÍES, M.E, PONTEVILLE, C¹

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

El objetivo general de este trabajo es identificar en las producciones escritas de los estudiantes de la asignatura Bioestadística, de las carreras de Farmacia y Bioquímica de la Universidad de Buenos Aires, las argumentaciones utilizadas en relación con el modelo de regresión lineal simple. Esto permitirá repensar la enseñanza del mencionado modelo, instaurando al momento de resolver situaciones problemáticas, la necesidad de conceptualizar en las clases prácticas, los elementos construidos a lo largo de la asignatura. Ya que, a lo largo de diversas instancias, observamos que los alumnos ponían más énfasis en la resolución vinculada a cálculos que en el reconocimiento de los conceptos presentes en el modelo.

Debido a los resultados obtenidos en el primer cuatrimestre de 2020 se tomó la decisión de solicitar explícitamente en los enunciados de los ejercicios: las definiciones de las variables, las especificaciones del modelo, los supuestos, los estadísticos que intervienen, las conclusiones correspondientes en términos del problema.

Los objetivos específicos fueron:

1) Realizar un análisis sistemático cualitativo, de las prácticas asociadas a los conceptos del modelo de regresión lineal simple, que aparecen en los exámenes. Para llevar a cabo este objetivo, se analizaron los exámenes correspondientes al segundo regulatorio del segundo cuatrimestre de 2020 (se analizaron 213 exámenes), que se desarrolló bajo entornos virtuales (tanto para el dictado de las clases como para la realización de los exámenes).

2) Comparar los resultados obtenidos en el cuatrimestre, anteriormente mencionado, con los del primer cuatrimestre de 2020 (se analizaron 226 exámenes), analizados previamente. Las diferencias identificadas se centraron esencialmente en la formulación de los ejercicios, ya que, en este segundo período se solicitó en forma explícita la definición de las variables aleatorias presentes, el enunciado del modelo, el planteo de los supuestos, las hipótesis, el estadístico de prueba y las fórmulas correspondientes.

Para responder a los objetivos planteados se realizó un análisis descriptivo de las variables intervinientes. Para estudiar el grado de asociación entre las variables se realizó una prueba de ji-cuadrado (objetivo 1). Para comparar los resultados obtenidos en el primer y segundo cuatrimestre de 2020 se aplicó la prueba de diferencia de proporciones (objetivo 2).

Esto reveló la necesidad de que se expliciten clara y minuciosamente las consignas de los ejercicios presentados en los exámenes, de manera de poder identificar el conocimiento respecto del modelo estudiado.

Palabras Clave: enseñanza, producciones escritas, entornos virtuales, regresión.

¹e-mail: myriam@ffyb.uba.ar

Diseño e implementación de un taller de divulgación de la estadística orientado a la escuela secundaria

Zangari C.A., Fernández V.M.¹, Lois-Milevicich J., Flaibani N., Fernández M.S.,
Pérez A.L., Miño M.H.²

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

En la sociedad científica actual existe cada vez mayor consenso en que el conocimiento no debe quedar dentro de las aulas y laboratorios universitarios, sino extenderse a la sociedad en general. En este marco, diseñamos un taller con el objetivo de promover la lectura crítica de información estadística proveniente de los medios de comunicación en estudiantes de nivel medio y proveer a los docentes de herramientas para abordar la temática en el aula. Además, se facilitó el acercamiento de estudiantes y docentes a los quehaceres de la ciencia, para favorecer el interés en la misma y mostrarles, de manera simple, lo que la estadística puede aportar a diferentes disciplinas científicas. El taller, llamado “*La estadística en la Biología, en las Ciencias y... ¡en la vida cotidiana!*”, formó parte del evento Semana de la Biología 2021, que anualmente ofrece la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires (Argentina) como actividad de divulgación científica. El mismo se ofreció en modalidad virtual a través de la plataforma Zoom y tuvo una duración de 90 minutos. Los asistentes fueron alumnos del último año de una escuela secundaria privada con orientación en Ciencias Naturales de Tierra del Fuego (Argentina), junto con sus tutores y docente. Se incluyeron cuatro actividades: 1) actividad de inicio destinada a indagar las ideas previas de los estudiantes sobre la estadística, mediante la construcción de una nube de palabras usando el software Mentimeter; 2) discusión acerca de la información estadística sesgada que frecuentemente se percibe en los medios de comunicación usando imágenes extraídas de diarios, internet y otros medios, como disparadores; 3) exposición breve de los talleristas sobre ejemplos de aplicación de la estadística en las investigaciones biológicas (extraídos de tesis de licenciatura) y 4) actividad de cierre indagando lo que más les interesó del taller, mediante el uso de notas adheridas a un Jamboard compartido. En relación a la actividad 1), en la nube de palabras aparecieron varias relacionadas con la disciplina (e.g. datos, variables, histograma), lo cual nos permitió saber que los estudiantes ya contaban con algunas nociones de estadística. A partir de esas palabras, se introdujo la actividad 2), en la cual los estudiantes participaron de manera escrita (a través del chat) y oral, dando su opinión sobre algunos gráficos con errores típicos (e.g. ausencia de escala en ejes, ejes que no inician en cero, gráficos de torta que muestran porcentajes totales mayores a 100) y maneras tendenciosas de reportar los datos (e.g. media o moda según el interés, generalizaciones a partir de muestras sesgadas). A raíz de la actividad 3, los estudiantes manifestaron su interés por conocer cómo organizar un trabajo de investigación, y explicaron que para recibirse debían realizar un trabajo final individual que incluyera la recolección y análisis de datos. Esto generó un diálogo que sirvió de cierre y devolución del taller en reemplazo del Jamboard. Sin embargo, la docente manifestó su interés por continuar trabajando sobre el Jamboard en las clases posteriores al taller. A través de las consultas de los estudiantes y comentarios de la docente, encontramos que esta escuela buscó un aporte extra para enriquecer algunos aspectos estadísticos del trabajo final de sus estudiantes. Por otra parte, creemos que encuentros como este donde se analiza críticamente la información estadística que recibimos cotidianamente, contribuyen a la formación de una ciudadanía más atenta.

Palabras Clave: Estadística descriptiva, herramientas virtuales, divulgación científica, modalidad taller, medios de comunicación.

¹e-mail: vfernandez@ege.fcen.uba.ar

²e-mail: marielaminio4@gmail.com

Tema 10: Espacial

Imputación de datos faltantes de la matriz de origen-destino en base a filtros espaciales.

Antonio Rey, Fernando Massa, Andrés Castrillejo y María Eugenia Riaño¹

Instituto de Estadística, Departamento de Métodos Cuantitativos, FCEA-UDELAR

Las encuestas de origen- destino, por lo general se realizan con el objetivo de estimar los flujos de transporte entre distintas regiones de una ciudad. Puede suceder, que si las unidades geográficas consideradas como regiones de transporte son pequeñas, el tamaño muestral no sea suficiente y se tengan celdas vacías en la matriz de origen- destino (OD). Si bien se entiende que el tamaño de estos flujos puede ser pequeño (ya que no fue captado por la encuesta), es necesario contar con una estimación para una adecuada planificación del transporte. Contar con una matriz OD completa es fundamental para el análisis de la movilidad urbana y de las necesidades de transporte de la población. Tradicionalmente los flujos de las matrices OD se estiman con modelos gravitacionales, herramienta que también se utiliza para la predicción de las celdas vacías de la matriz OD.

El objetivo del presente trabajo es comparar el desempeño de los modelos gravitacionales que incorporan la autocorrelación espacial en la imputación de datos faltantes de la matriz OD, con los modelos que no consideran la autocorrelación espacial. La autocorrelación espacial se incorpora utilizando filtros espaciales, obtenidos del conjunto de vectores propios de una apropiada transformación de la matriz de pesos espaciales.

El filtrado espacial parte del supuesto de que la correlación espacial existe debido a una mala especificación del modelo, es decir, existe un conjunto de variables omitidas, que hacen que los errores del modelo presenten autocorrelación espacial. Obteniendo los vectores propios de una transformación de la matriz de pesos espaciales, se obtiene un conjunto de variables “proxy” que al ser incorporadas al modelo, hacen que los residuos del modelo no presenten autocorrelación espacial. La incorporación de los vectores propios se realiza mediante un proceso similar al “forward”, incorporando vectores hasta que la autocorrelación espacial medida con el Índice de Moran, sea eliminada de los residuos del modelo.

Para la evaluación de desempeño se cuenta con la matriz origen-destino completa para la ciudad de Bogotá obtenida de la Encuesta de Movilidad de Bogotá realizada en el año 2015. Se simulan distintas cantidades de datos faltantes, teniendo en cuenta el tamaño del flujo original y aumentando la cantidad de celdas vacías de la matriz.

Tomando a la Raíz del Error Cuadrático Medio de la predicción como medida de desempeño, los resultados demuestran una ganancia de hasta el 71 % cuando se trata de flujos pequeños (el escenario más probable), demostrando que la importancia de incorporar de la autocorrelación espacial en la predicción de los flujos faltantes de la matriz.

Palabras Clave: Estadística espacial, filtrado espacial, imputación, matrices Origen-Destino

¹e-mail: arey\$_1986@yahoo.com, fmassa@iesta.edu.uy, castrillejoandres@gmail.com, eugenia@iesta.edu.uy

Análisis de Componentes Principales con datos georeferenciados. Una aplicación en la construcción de un Índice de Calidad de Vida para la Argentina

García, Fernando; Stimolo María Inés; Ferrer Estanislao¹

Instituto de Estadística y Demografía, Facultad de Ciencias Económicas, Universidad Nacional de Córdoba. Argentina

El índice de Calidad de Vida (ICV) en Argentina fue construido por un equipo de expertos dirigido por G.A Velázquez con el objetivo de conocer las diferencias sociales, económicas, ambiental y regional de la calidad de vida de la población argentina, para lo cual se han considerado principalmente dos dimensiones: la dimensión socioeconómica (educación, salud y vivienda) y la ambiental (problemas ambientales, recursos recreativos naturales y recursos recreativos socialmente construidos).

Este indicador considera la información a nivel departamental, por lo que resulta un valioso instrumento de gestión de política públicas de salud, vivienda, educación, medio ambiente y cultura. De esta manera resulta una medida que cuantifica un nivel de vida de la población fuera de la percepción de los distintos actores sociales.

En este trabajo se propone un análisis de las ponderaciones de los indicadores que conforman cada dimensión definida en el ICV. El análisis de Componentes Principales (ACP) es una herramienta muy utilizada para analizar como participa cada indicador en la estructura de variabilidad de los datos. Sin embargo, como los indicadores están georeferenciados la estructura de co-variación reflejada por un análisis multivariado clásico puede verse afectada por los patrones espaciales subyacentes en los datos. En tal sentido, resulta adecuado aplicar el método conocido como MULTISPATI, el cual se basa en el ACP pero incorpora la restricción dada por los datos espaciales mediante el cálculo del índice de Moran para medir la dependencia o correlación espacial entre las observaciones. A partir de la agregación (a través de una suma ponderada) de los valores de las componentes espaciales (CPs) se definió el índice, utilizando como ponderación los valores propios asociados a cada componente.

Los resultados muestran que las CPs seleccionadas permitirían una visualización mejor de la variabilidad espacial. De esta manera, sería recomendable avanzar en la construcción del Índice de calidad de vida a partir de las CPs obtenidas a través de MULTISPATI.

Palabras claves: Calidad de Vida, Datos georeferenciados, Índice, MULTISPATI. **Palabras Clave:**

¹e-mail: fernando.garcia@unc.edu.ar, maria.ines.stimolounc.edu.ar, estanislao.ferrer@mi.unc.edu.ar

Una aplicación de la estadística espacial en el análisis de feminicidios en la Ciudad de México México

Víctor Manuel Ulloa Arellano, Verónica del Carmen Quijada Monroy¹

Facultad de Estudios Superiores Acatlán, Posgrado en Economía y Centro Tecnológico para la Educación a Distancia, Universidad Nacional Autónoma de México

En la República Mexicana, el artículo 21 de la Ley General de Acceso de las Mujeres a una Vida Libre de Violencia, define al feminicidio como «la forma extrema de violencia de género contra las mujeres, conformada por el conjunto de conductas misóginas que pueden conllevar impunidad social y del Estado y puede culminar en homicidio y otras formas de muerte violenta de mujeres». Ante la gravedad de este problema, es indispensable su estudio a través de métodos rigurosos tales como son las técnicas estadísticas y de ciencia de datos, particularmente la estadística espacial y la minería de datos. Si bien el Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública concentra las estadísticas de las fiscalías de las 32 entidades del país relativas a este delito, los datos que se generan ante delitos violentos contra las mujeres no resultan, en general, como elementos suficientes para un análisis, visibilización y toma de decisiones efectivo, ya que únicamente la Ciudad de México a través de su portal de datos abiertos, incorpora entre otras variables, las coordenadas geográficas relativas a este tipo de delito. En este trabajo se presenta una aplicación de estadística espacial implantada en RStudio, como punto de partida para establecer un marco de trabajo que permita la construcción de mapas de violencia contra las mujeres para la detección de patrones y contribuir al estudio especializado de este fenómeno.

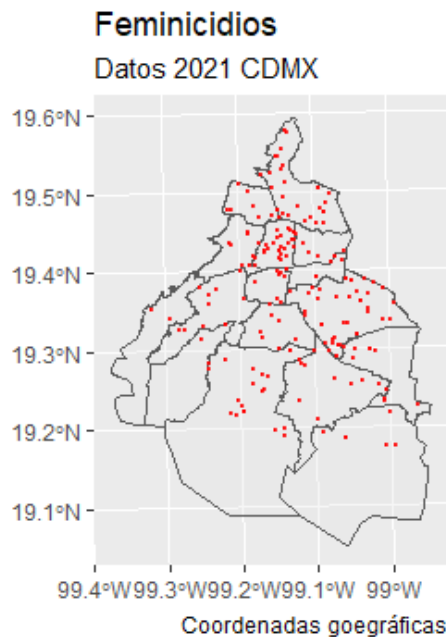


Figura 1: Georreferenciación de feminicidios en CDMX. Fuente: elaboración propia.

Palabras Clave: estadística espacial, feminicidio, datos abiertos, georreferenciación, RStudio

¹e-mail: victormu@acatlan.unam.mx, quijada@acatlan.unam.mx

Tema 11: Análisis Longitudinal

Análisis de datos longitudinales, tres modelos que llevan al mismo resultado

Mazzucco, Andrea Carolina, Sfer, Ana María¹

Departamento de Matemática, Facultad de Ciencias Exactas y Tecnología, Universidad Nacional de Tucumán

El objetivo del presente trabajo, es estudiar, aplicar y comparar las técnicas Análisis Multivariado de la Varianza (MANOVA) para medidas repetidas, Ecuaciones de Estimación Generalizadas (GEE) y Análisis de Modelos de Coeficientes Aleatorios o Modelos Mixtos para un estudio de datos longitudinales y observar como ellas llevan a resultados comparables, consistentes y en algunos casos idénticos .

Se utiliza una base de datos reales: Datos del Ingreso a la Facultad de Ciencias Exactas y Tecnología de la Universidad Nacional de Tucumán. El objetivo es determinar la relación entre el tipo de ingreso que realizó el alumno y su efecto sobre el avance en la carrera. Se considera la cohorte 2009 y su evolución durante 10 años. Se utilizan los paquetes estadísticos STATA y SPSS. Los resultados indican que el porcentaje de avance disminuye con el tiempo y que los alumnos con mejor desempeño, es decir con mayor porcentaje de avance, son aquellos que aprobaron el ingreso mediante una prueba de suficiencia. Las variables estudiadas son: Legajo, Tipo de ingreso (curso agosto-diciembre 2008; curso febrero 2009; examen de suficiencia y sin ingreso), Tiempo y Porcentaje de avance.

El ajuste MANOVA con medidas repetidas indica que, el tiempo y el tipo de ingreso, son estadísticamente significativos, es decir, que el avance promedio del alumno en la carrera no es el mismo en los 10 momentos estudiados y también cambia con al tipo de ingreso. Para entender cuál es el comportamiento de la respuesta con respecto a estas variables es necesario una representación gráfica del porcentaje de avance a lo largo del tiempo. Los ajustes correspondientes a GEE y Modelos Mixtos indican los mismos resultados, es decir, que nuevamente el tiempo y el tipo de ingreso son estadísticamente significativos. Los coeficientes se muestran en el Cuadro 1. En estos ajustes si se puede observar a partir del signo del coeficiente que el

	GEE			Modelos Mixtos		
	Coficiente	Desv. Est.	Valor p	Coficiente	Desv. Est.	Valor p
Tiempo	-0,0033	0,0003	< 0,001	-0,0033	0,0003	< 0,001
Curso Ago-Dic 2008	0,0316	0,0077	< 0,001	0,0316	0,0082	< 0,001
Curso Feb 2009	0,0158	0,0075	0,035	0,0158	0,0079	0,045
Suficiencia	0,0317	0,0095	0,001	0,0318	0,0100	0,001
Constante	6,6912	0,5660	< 0,001	6,6318	0,5501	< 0,001

Cuadro 1: Coeficientes de Modelo GEE y Modelos mixtos con estructura de correlación intercambiable por tiempo y tipo de ingreso

porcentaje de avance disminuye con el tiempo y que se incrementa en mayor medida cuando el alumno ingresa mediante examen de suficiencia.

Con respecto al diagnóstico del modelo solo se verificó que los errores $\varepsilon_{jit} \sim N(0, \sigma^2)$

Al cabo de aplicar las diferentes herramientas para analizar datos longitudinales se concluye que los resultados son consistentes entre MANOVA con medidas repetidas y los modelos GEE y Modelo Mixto. Los resultados son idénticos entre el modelo GEE y Modelo Mixto con ordenada al origen aleatoria respetando la misma estructura de correlación. La decisión de qué herramienta aplicar dependerá del objetivo de estudio y de la presencia o no de datos faltantes. Si sólo se quiere estudiar si ciertas variables predictoras cambian o no la respuesta se utiliza MANOVA con medidas repetidas. Si se quiere explicar o modelar la relación entre las variables predictoras y la respuesta se utilizan modelos GEE o Modelos mixtos.

Palabras Clave: datos longitudinales, correlacionados, porcentaje de avance

¹e-mail:acmazzucco@herrera.unt.edu.ar, asfer@herrera.unt.edu.ar

Comparación de Modelos Mixtos Para Describir el Crecimiento Físico en la Población Infantil Vulnerable en Argentina

Tomás González Garello¹, M. Soledad Fernández, Pablo Turjanski, Agustín Dramis,
Adriana Pérez, Gerardo Cueto

*Grupo de Bioestadística Aplicada, Instituto de Ecología, Genética y Evolución de Buenos Aires;
Instituto de Cálculo; Instituto de Ciencias de la Computación, FCEN, UBA-CONICET*

Objetivo: Seleccionar el modelo estadístico de crecimiento con mejor ajuste a registros longitudinales de talla y perímetro cefálico durante los primeros años de vida en la población infantil argentina con cobertura exclusiva del sistema público de salud

Población de Estudio: Utilizamos registros antropométricos de niños y niñas beneficiarios/as del programa SUMAR del Ministerio de Salud, con al menos 10 mediciones de talla entre los 0 y 5 años (61.946 individuos femeninos y 61.315 individuos masculinos, con un total de 801.528 y 793.439 registros, respectivamente), y con al menos 5 mediciones de perímetro cefálico durante el primer año de vida (136.846 individuos femeninos y 139.304 individuos masculinos, con un total de 814.945 y 830.084 registros, respectivamente)

Metodología: Comparamos el ajuste y poder predictivo de 5 modelos estructurados mixtos (Count, Berkey-Reeds, Gompertz Adaptado, JenSS Bayley y JenSS Bayley Adaptado), sobre el crecimiento físico de individuos de 0 a 5 años en talla y del primer año de vida en perímetro cefálico. Para cada medida antropométrica y sexo, ajustamos cada modelo a los registros correspondientes al 70 % de los individuos elegidos al azar (base de entrenamiento). Comparamos el buen ajuste de cada modelo utilizando el criterio de información de Akaike (AIC), mientras que comparamos el poder predictivo de cada modelo a través de la raíz del error cuadrático medio (RMSE) obtenido con el 30 % de los restantes registros (base de prueba). A partir del mejor modelo, según sexo y medida antropométrica, ajustamos la curva de crecimiento medio en función de la edad, a la vez que obtuvimos los indicadores individuales de crecimiento

Resultados: Los 5 modelos convergieron para las 2 medidas antropométricas de ambos sexos. No observamos patrones anómalos en el comportamiento de los residuos en función de la edad. Para los registros de talla, el mejor modelo por buen ajuste resultó ser -para ambos sexos- el modelo de JenSS Bayley Adaptado, mientras que para los registros de perímetro cefálico el mejor modelo por buen ajuste resultó ser -para ambos sexos- el modelo de Count. Para cada medida antropométrica y sexo, todos los modelos mostraron un poder predictivo similar, con un RMSE de predicción que varió entre 3,56 y 3,66 cm en los modelos de talla, y entre 1,55 y 1,61 cm en los modelos de perímetro cefálico

Conclusión: El modelo JenSS Bayley Adaptado es el que mejor describe el crecimiento medio y el crecimiento individual en talla para el rango de 0 a 5 años, para ambos sexos. El modelo Count es el que mejor describe el crecimiento medio y el crecimiento individual en perímetro cefálico durante el primer año de vida, para ambos sexos. Para cada medida antropométrica y sexo, todos los modelos resultaron similares para predecir la talla o el perímetro cefálico de la base de prueba.

Validar el modelo que mejor describa el crecimiento físico en la infancia según distintas medidas antropométricas es un paso fundamental para diseñar estudios de investigación en factores de riesgo en el crecimiento infantil en Argentina

Palabras Clave: Crecimiento Físico Infantil, Modelos Mixtos, Talla, Perímetro Cefálico

¹e-mail: tomas21.gg@gmail.com, autor2@mail.com

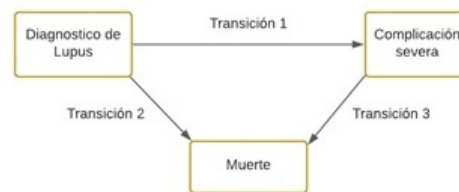
Estimación de probabilidades de transición mediante modelos de estados múltiples para el estudio del lupus

Arnesi, Nora E., Boggio, Gabriela S., Harvey Guillermina B., Settecase Eugenia¹

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística, Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario

Al estudiar el curso de una enfermedad puede resultar de interés considerar no sólo el estado inicial de los pacientes al momento del diagnóstico y el estado al final del seguimiento, sino también estados intermedios por los que puede transitar cada paciente. Los denominados modelos de estados múltiples permiten analizar la experiencia de supervivencia teniendo en cuenta estas secuencias o trayectorias diferentes. El objetivo de este trabajo es estimar probabilidades de transición, condicionando la ocurrencia de ciertos eventos a trayectorias particulares de pacientes con diagnóstico de lupus mediante el ajuste de modelos de estados múltiples. Para ello se considera un esquema simple que involucra tres estados: diagnóstico de lupus, complicación severa –renal y/o cardiovascular- y muerte (Figura 1).

Figura 1: Esquema multiestado para el lupus



A partir del esquema considerado se identifica la respectiva matriz de transición. En cuanto a la información disponible, fue necesario reestructurar la base de datos pasando del formato de un registro por paciente a uno con una fila por cada transición posible para cada paciente. Con respecto a la especificación de la escala del tiempo, se elige el enfoque denominado “reloj hacia adelante” asumiendo la propiedad markoviana, la cual establece que el estado futuro del paciente depende de la historia sólo a través del presente. Ello permite la estimación de las probabilidades de transición a través de un proceso de Markov no homogéneo en el tiempo. La formulación del modelo de estados múltiples como proceso de conteo permite su ajuste mediante un modelo de hazards proporcionales estratificado, donde los estratos se corresponden con una partición del tiempo de seguimiento de los pacientes de acuerdo a las transiciones entre los estados por los que atraviesan. El modelo considerado asume diferentes efectos de las covariables para las transiciones. A partir de las hazards basales estimadas para cada transición se pudieron obtener las hazards acumuladas para perfiles de interés y en función de ellas, mediante el método de Aalen-Johansen, las estimaciones de las respectivas probabilidades de transición para predecir eventos futuros en diferentes momentos. Así, por ejemplo, para mujeres jóvenes de etnia blanca se pudo observar que la probabilidad de presentar una complicación severa aumenta notablemente en los primeros meses pos-diagnóstico y luego el crecimiento es más lento estabilizándose en un valor cercano a 0,4 a los 5 años. En cambio, estas probabilidades de transición se reducen notablemente si transcurrido un año, las pacientes permanecen en el estado inicial. Se continuará explorando el uso tanto de los modelos de Markov como así también semi-Markov en el análisis de datos de supervivencia ante esquemas de estados múltiples más complejos que el considerado en esta ocasión.

Palabras Clave: tiempos de supervivencia; modelo de Markov; modelo de Cox.

¹e-mail: narnesi@fcecon.unr.edu.ar, gboggio@fcecon.unr.edu.ar, gharvey@fcecon.unr.edu.ar, eugenia.settecase@gmail.com

A joint spatial modeling for mosquito-borne diseases in Brazil

Jessica Pavani^{a1}, Leonardo Bastos^b, and Paula Moraga^c

^a*Department of Statistics, Pontificia Universidad Católica de Chile, Santiago, Chile*

^b*Scientific Computing Program, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil*

^c*Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*

Mosquitoes are responsible for transmission of a great part of infectious diseases and lead to hundreds of thousands of deaths annually all over the world. In the Americas, diseases as dengue and chikungunya co-circulate, causing great damage to the population. In 2020, for instance, more than 2 million cases were reported on the continent, being Brazil the responsible for most of them. Although they are transmitted by the same mosquito, it remains unclear whether there exists a relationship between both diseases. In this study, we analyse the geographic distributions of dengue and chikungunya over the years 2016 to 2020 in the Brazilian state of Ceará. To pursue this, we use a Bayesian hierarchical spatial approach to jointly analyse two arboviruses. This modelling includes spatial covariates as well as specific and shared spatial effects that take into account the potential autocorrelation between diseases. Our findings allow us to identify areas with high risk of one or both diseases, and show that only 4% of the areas present high relative risk for both diseases, suggesting a competition between viruses. This study advances the understanding of the geographic patterns and the identification of risk factors of dengue and chikungunya, and can help health decision-making.

Keywords: Bayesian inference, chikungunya fever, dengue fever, disease mapping, spatial modeling, INLA.

¹e-mail: jpgavani@mat.uc.cl

Clasificación en modelos mixtos no lineales semiparamétricos usando P-Splines y el algoritmo SAEM

Maritza Márquez, ^{a 1} Cristian Meza, ^{a 2} Dae-Jin Lee ^{b 3} y Rolando De la Cruz ^{c 4}

^a CIMFAV, Facultad de Ingeniería, Universidad de Valparaíso, Valparaíso, Chile

^b BCAM, Basque Centre for Applied Mathematics, Bilbao, País Vasco, España

^c Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibañez, Santiago, Chile

Marshall y Barón (2000), Arriba-Gil, et al. (2015) y De la Cruz, et al. (2017), analizaron datos provenientes de un estudio clínico sobre el riesgo de pérdida para un grupo de mujeres chilenas embarazadas. Se clasificaron estas mujeres en dos grupos: las que tuvieron un embarazo que llegó a término, grupo “normal”, y las que tuvieron una pérdida, grupo “anormal”. En particular, estos autores modelaron la concentración de la hormona beta-HCG en 173 mujeres durante el primer trimestre usando modelos mixtos, de las cuales 124 tuvieron un embarazo normal y las 49 restantes un embarazo anormal, como apreciamos en la Figura 1. En nuestro trabajo de investigación, extendemos estos trabajos previos, usando un modelo semiparamétrico de efectos mixtos no lineal utilizando P-Splines y el algoritmo SAEM (Delyon, et al. 1999). Adicionalmente, logramos llevar a cabo la clasificación de un modelo mixto no lineal (paramétrico y semiparamétrico) vía Importance Sampling (IS).

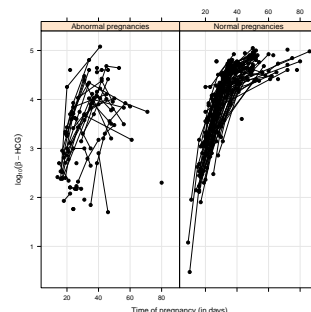


Figura 1: Perfiles observados de $\log_{10}(\beta - HCG)$ para el grupo anormal (panel izquierdo) y el grupo normal (panel derecho).

Para efectos de nuestra investigación proponemos: **a)** un modelo mixto no lineal logístico con tres efectos aleatorios ($NLME_{(3)}$), a diferencia de los trabajos previos (Marshall y Barón, 2000) que solo usaba un efecto aleatorio en el numerador ($NLME_{(1)}$) y **b)** un modelo aditivo de efectos mixtos no lineales con penalización spline ($SMNLME$). Ambos modelos fueron estimados utilizando el algoritmo SAEM, que resulta ser una aproximación estocástica del algoritmo EM. Basados en estos modelos y buscando minimizar la probabilidad de clasificación errónea, indicando el grupo al que pertenece cada individuo. En la Tabla 1 presentamos la matriz de confusión para cada uno de los modelos, incluido el $NLME_{(1)}$ propuesto por Marshall y Barón en el 2000, para efectos de comparación.

Grupo	$NLME_{(1)}$		$NLME_{(3)}$		SMNLME		Total
	Normal	Anormal	Normal	Anormal	Normal	Anormal	
							Total (173)
Normal	118	6	120	4	122	2	124
Anormal	21	28	20	29	20	29	49
AUC $NLME_{(1)}$	0,859						
AUC $NLME_{(3)}$	0,870						
AUC SMNLME	0,884						

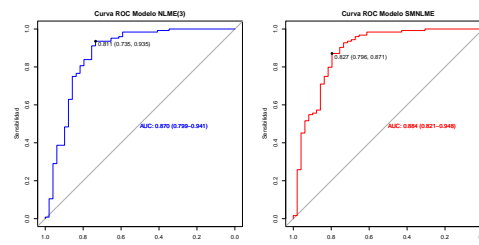


Tabla 1: Resultados de la clasificación de mujeres embarazadas usando Importance Sampling para los modelos $NLME_{(3)}$ y $SMNLME$ implementados bajo el algoritmo SAEM. Esto comparado con el modelo $NLME_{(1)}$.

Figura 2: Comparación de las curvas ROC y los AUC de los modelos $NLME_{(3)}$ y $SMNLME$ usando el método de Importang Sampling.

Por lo tanto, logramos apreciar que el modelo $SMNLME$ presenta mejoras tanto en la clasificación (ver Tabla 1) como en el AUC (ver Figura 2), con respecto a los otros modelos. Sin embargo la mejora se limita al grupo normal. De acuerdo a lo expuesto y para finalizar, es importante destacar que se logró proponer la clasificación para un modelo mixto no lineal vía IS e implementando el algoritmo SAEM, además se extendió una versión semiparamétrica del modelo $NLME_{(3)}$ obteniendo mejoras en la matriz de confusión y su AUC.

Palabras Clave: Datos longitudinales, clasificación supervisada, modelo mixto no lineal, método P-spline, algoritmo SAEM, Importance Sampling.

¹e-mail: maritza.marquez@postgrado.uv.cl

²e-mail: cristian.meza@uv.cl

³e-mail: dlee@bcamath.org

⁴e-mail: rolando.delacruz@uai.cl

Alternative subgroup joint analysis proposal of nonlinear longitudinal and time-to-event data for modeling pregnancy miscarriage

Cristian Meza^{*1}, Rolando de la Cruz[†], Marc Lavielle^{††}, Vicente Núñez-Antón^{**}

^{*}INGEMAT-CIMFAV, Facultad de Ingeniería, Universidad de Valparaíso

[†]Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez

^{††}CMAP, Ecole Polytechnique, CNRS, Institut Polytechnique de Paris

^{**}Department of Quantitative Methods, University of the Basque Country UPV/EHU

Pregnancies achieved through *in-vitro* fertilization (IVF) are associated with adverse first trimester outcomes in comparison to spontaneously achieved pregnancies. Human chorionic gonadotrophin β subunit (β -HCG) is a well-known and accurate biomarker for the diagnosis and monitoring of pregnancy after IVF. Low levels of β -HCG during the first trimester of pregnancy are related to miscarriage, ectopic pregnancy and failure of IVF procedure. Longitudinal profiles of β -HCG can be utilized to distinguish between normal and abnormal pregnancies, and to assist and guide the clinician in better management and monitoring of post-IVF pregnancies. Therefore, being able to assess the association between longitudinally measured β -HCG and time to early miscarriage is of crucial interest to clinicians. A common joint modeling approach proposal to achieve this objective is to use subject-specific random effects in a mixed effects model for longitudinal β -HCG data as predictors in a model for the time-to-event (TTE) data. This work was motivated by an observational study with normal and abnormal pregnancies where concentrations of β -HCG were measured in 173 young women during a gestational age of 9-86 days at a private clinic in Santiago, Chile. Some women experienced a miscarriage but their exact event times were unknown, in such case we have interval censored data, assuming that the event occurred between the last time of the observed β -HCG measurement and ten days after. We consider a nonlinear mixed effects (NLME) model for both normal and abnormal pregnancies, but the joint model is considered only for the subgroup of miscarriage women. All the estimation procedures are based on the Stochastic Approximation of the EM (SAEM) algorithm implemented in the Monolix software. Our analyses showed that convergence of this algorithm here is fast, specially when compared to alternative likelihood approaches for which, in some cases, convergence was not achieved.

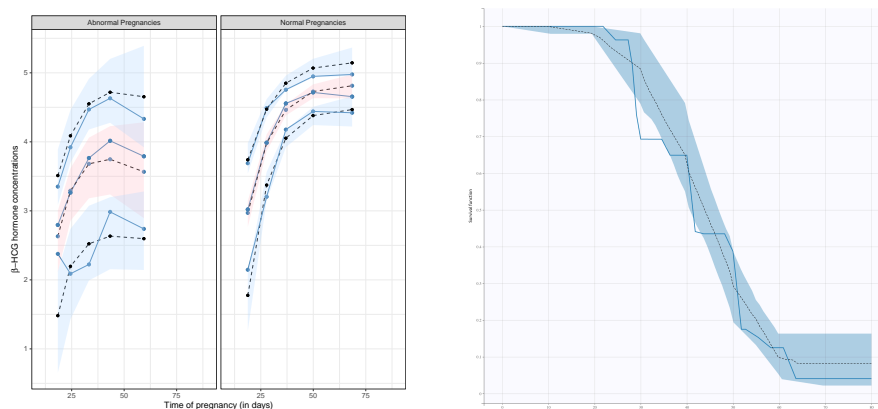


Figura 1: **Left:** Visual predictive check for the β -HCG hormone concentrations with the 95% prediction intervals for the 10th, 50th and 90th percentiles, empirical percentiles (in solid line) and predicted percentiles (in dashed line); **Right:** Survival function estimated for abnormal group using the SAEM algorithm. In blue, the predicted interval for the Kaplan Meier plot, in dashed line the predicted median.

Palabras Clave: Joint model; longitudinal data; time-to-event; SAEM algorithm.

¹e-mail: cristian.meza@uv.cl, rolando.delacruz@uai.cl, Marc.Lavielle@inria.fr, vicente.nunezanton@ehu.eus

Transiciones hacia la dependencia con el envejecimiento en Europa

Alejandra Marroig¹

Instituto de Estadística, Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay

Durante el envejecimiento de las personas pueden aparecer limitaciones funcionales en donde podría ser necesaria la ayuda de otras personas para realizar actividades diarias. Algunas personas pueden tener discapacidad para la vida diaria y necesidad de cuidados en una etapa más temprana de la vida o con mayor urgencia, mientras que otras pueden conservar su independencia durante más tiempo.

En este proyecto analizo las transiciones hacia la discapacidad y dependencia de adultos mayores. Utilizo datos desde 2004 hasta 2013 de diez países de la *Survey of Health, Ageing and Retirement in Europe* (SHARE). Los estados de discapacidad y dependencia se definen a partir de las limitaciones en las Actividades Básicas de la Vida Diaria (ABVD) y Actividades Instrumentales de la Vida Diaria (AIVD) para los participantes de 65 años o más al inicio del estudio (n=20128). Ajusto modelos de sobrevivencia de múltiples estados para datos longitudinales, también llamados modelos de markov para datos longitudinales. En ellos evalué el rol de la edad, el sexo y la educación en las transiciones entre estados de dependencia, incluido el fallecimiento, en los diferentes países europeos. También obtengo la esperanza de vida libre de discapacidad y dependencia para cada país.

Los resultados para todos los países agrupados, muestran que las transiciones hacia la discapacidad y la dependencia varían con la edad, el sexo y la educación. Por su parte, en los análisis por país, la probabilidad de transición a la discapacidad y dependencia aumenta hasta los 70 años, resultado consistente entre países. También se evidencia el rol de la edad y el sexo en el paso hacia la discapacidad y dependencia aún considerando el fallecimiento, sin embargo emergen diferencias entre países. En la mayoría de los países, las mujeres tienen mayor riesgo de transición a la discapacidad que los hombres, considerando también la transición a la muerte. Sin embargo, sólo en unos pocos países este riesgo sigue siendo mayor cuando se utiliza la medida "proxy" de la dependencia. En la mayoría de los países, los resultados en términos de las esperanzas de vida sugieren que las mujeres viven más años con dificultades y pueden necesitar la ayuda diaria de otras personas durante más tiempo que los hombres.

Con este trabajo se puede propiciar una detección de la dependencia en una población que envejece que podría informar sobre el diseño de estrategias de prevención y políticas de cuidado, que consideren las especificidades para hombres y mujeres y mejoren la calidad de vida de los adultos mayores y sus familias.

Palabras Clave: Envejecimiento, Actividades de la Vida Diaria, Dependencia, Modelo de Múltiples Estados, Europa

¹e-mail: alejandra.marroig@fcea.edu.uy

Tema 12: Machine Learning

Detección de puntos de cambios en una caminata aleatoria mínima con aprendizaje

V́ctor Hugo V́zquez Guevara¹, Ram3n ́lvarez-Vaz²

Facultad de Ciencias F́sico Matemáticas, Benemérita Universidad Autónoma de Puebla, México
Instituto de Estadística, Facultad de Ciencias Econ3micas y de Administraci3n, Universidad de la República, Uruguay

En este trabajo se presenta una caminata aleatoria propuesta en 2014 por U. Harbola, N. Kumar y K. Lindenberg en el art́culo *Memory-induced anomalous dynamics in a minimal random walk model. Phys. Rev. E 90, 022136* que generaliza a la caminata aleatoria del elefante que vio la luz en 2004.

En este modelo discreto, las part́culas tienen dos posibilidades en cada instante: pueden moverse hacia adelante o permanecer quietas con ciertas probabilidades condicionales (p y q respectivamente). Esta caminata aleatoria es no markoviana y tiene memoria completa de su pasado.

La caminata aleatoria ḿnima, cuyo movimiento y posici3n en el instante $n \geq 1$ son denotados por X_n y S_n respectivamente evoluciona como sigue:

1. En el instante $n = 0$ una part́cula que sigue este modelo se encuentra en el origen, es decir, $S_0 = 0$.
2. En $n = 1$, la part́cula decide moverse una unidad a la derecha con probabilidad $s \in (0, 1)$ o permanecer en el origen con probabilidad $1 - s$. Entonces $S_1 = X_1$ tiene distribuci3n Bernoulli con parámetro s
3. En cualquier instante $n \geq 2$, la part́cula elige un instante del pasado de forma uniforme, por ejemplo $1 \leq k(n) \leq n - 1$. Entonces, su siguiente movimiento X_n , será condicionado por el valor observado de $X_{k(n)}$ de la siguiente manera:

Por un lado, si $X_{k(n)} = 0$ entonces X_n tiene distribuci3n Bernoulli con parámetro $q \in [0, 1]$. Mientras que si $X_{k(n)} = 1$, la distribuci3n condicional de X_n también es Bernoulli pero con parámetro $p \in [0, 1]$

4. La posici3n de la part́cula al instante $n \geq 1$ es tal que

$$S_n = S_{n-1} + X_n, \tag{1}$$

con $S_0 = 0$ y $X_n \in \{0, 1\}$.

Sin embargo, en este trabajo se propone un enfoque en el cual es posible que los parámetros p y q evolucionen con el tiempo (de forma no decreciente) con cierta probabilidad en determinados instantes que pueden ser aleatorios. Bajo este esquema, si se contara con una trayectoria de un proceso de caminata aleatoria ḿnima con parámetros no estáticos, puede resultar de interés el determinar los instantes en los que la evoluci3n de los parámetros sea evidenciada, y para esto se explorarán herramientas provistas por el software Mat-Lab en complemento con ejemplos numéricos.

Palabras Clave: Machine Learning, caminata aleatoria, puntos de cambio

¹e-mail: vvazquez@fcfm.buap.mx

²e-mail:ramon.alvarez@fcea.edu.uy

Clasificación mediante K-Medias de un área sembrada

Patricia Girimonte, Javier I. García Fronti¹

FCE (CIMBAGE), FFyB (Cátedra de Matemática), Universidad de Buenos Aires

La actividad agrícola es una de las actividades económicas más importantes de nuestro país, lo cual impulsa el desarrollo de estrategias para maximizar la productividad de los cultivos, reduciendo los costos y el impacto ambiental.

Una herramienta utilizada con este objetivo es la teledetección satelital, cuya aplicación comienza en la década del 70, donde también se definieron los primeros índices relacionados con la productividad, siendo uno de los más utilizados el Índice de Vegetación de Diferencia Normalizada (NDVI) también llamado Índice Verde (Rouse, et al, 1973).

A partir de la política de Datos Abiertos iniciadas por la NASA (National Aeronautics and Space Administration) y la ESA (European Spatial Aeronautics) en los años 2013 y 2017 respectivamente, las imágenes satelitales se encuentran disponibles más allá de las áreas gubernamentales, representando una oportunidad para generar conocimiento, tanto en el ámbito académico, como en el sector público y privado.

En este trabajo se seleccionó una imagen satelital del Landsat 8 de la NASA del 12 de marzo de 2019 de la región pampeana norte, donde se encuentra uno de los departamentos de mayor producción de soja. A partir de la imagen obtenida se seleccionó un polígono de interés donde se calculó el NDVI para cada píxel, con el objeto de clasificar el área seleccionada mediante el método de aprendizaje automático de clasificación no supervisado de K-Medias.

Una vez obtenidas las 11 bandas de la imagen se realizó su procesamiento en R 3.6.1 con las librerías raster (Hijmans, 2019), dplyr (Wickham et.al. ,2019), sp (Pebesma, E.J., R.S. Bivand, 2005) y rgdal (Bivand, et.al, 2019). Como nuestro objeto es clasificar la cobertura del suelo a partir de NDVI, el análisis se realizó con el apilado de las bandas 4 (infrarrojo cercano),5 (infrarrojo de onda corta) y 3 (rojo), dado que permite tener una apreciación visual de las áreas cultivadas.

Para clasificar la zona seleccionada mediante el algoritmo de K-medias, se trabajó con el total de los píxeles. Se utilizó la función kmeans de R (R Core Team, 2019) considerando un máximo de 500 interacciones, y el algoritmo de Lloyd (Lloyd, 1982). Para determinar el número de clusters a considerar se aplicó el método heurístico “del codo”. Con los resultados obtenidos mediante el método de K-medias considerando de 1 a 15 clusters, comparando los resultados obtenidos, resultó adecuado para la región analizada considerar 9 clusters.

A partir del análisis realizado, para la región de interés, donde tenemos un conocimiento previo de que se trata de zonas con plantaciones de soja mayoritariamente, hemos obtenido como era esperable valores de NDVI mayores a 0.5 mayoritariamente.

Palabras Clave: K-medias, teledetección, NDVI, productividad

¹e-mail: patriciagirimonte@gmail.com, javier.garciafronti@economicas.uba.ar

Estudio de la resistencia a la fractura de polipropileno iPP mediante modelos de regresión MARS

Luis Arenas¹, Guillermo Sabino¹, Edgardo Santarelli²

¹Depto.de Estadística, Fac.de Economía y Administración, ²Centro de Ciencia y Tecnología de Materiales, Fac.de Ingeniería, Universidad Nacional del Comahue, Neuquén, Argentina

Introducción: El polipropileno isotáctico (iPP) es un material polimérico que presenta un comportamiento de transición que genera dispersión de resultados de la tenacidad a la fractura medida, lo que dio origen a modelos de tratamiento estadístico. Algunos modelos son inconsistentes con el comportamiento que presenta el material por lo cual pierden aplicabilidad.

En este trabajo se analiza un conjunto de resultados de resistencia a la fractura obtenidos sobre probetas de iPP, variando tamaños y velocidades de ensayo. La estimación de la función media de la tenacidad se efectúa con modelos MARS (Multivariate Adaptive Regression Spline). MARS es una técnica de regresión no paramétrica que combina regresión mediante splines y métodos de selección de modelos. Construye las funciones base spline de manera adaptativa mediante la determinación de los valores nodos apropiados para las diferentes variables y obtiene modelos reducidos aplicando técnicas de selección de variables. La función de variancia se estima mediante modelos de regresión lineal y con la utilización combinada de ambos tipos de modelos se obtienen intervalos de predicción. Empleando los límites inferiores de los intervalos de predicción es posible la estimación de valores de la cola izquierda de la distribución de la tenacidad.

Materiales y métodos: Se utilizaron cuatro espesores de probeta (5, 10, 15 y 25mm). Los efectos de velocidad sobre el comportamiento plástico del material se determinaron mediante ensayos de tracción a cuatro velocidades (0,014; 0,08; 0,13 y 0,195 (1/seg)), resultando así entre 10 y 12 réplicas por celda en un experimento factorial con un total de 180 probetas ensayadas.

El análisis y la obtención de los modelos MARS se realizó en R (ver 4.1.0) con los paquetes *caret* (6.0.88) y *earth* (5.3.1). El paquete *earth* permite estimar simultáneamente tanto la función media como la función de variancia. La variable de respuesta considerada fue la resistencia a la fractura (J_c). Las variables predictoras fueron el espesor de la probeta (B), la velocidad de deformación (V_{def}), el crecimiento previo de fisura (Δa), la restricción a la deformación plástica (α), el desplazamiento en carga máxima (dP_{max}) y la tensión de fluencia (s_y). Los datos se partitionaron de manera estratificada y aleatoria para obtener el conjunto de entrenamiento y de test. En el primero, se repitió cincuenta veces la validación cruzada 10-folds. Se determinó de este modo: la cantidad de funciones base y el grado máximo de interacción entre predictores. A partir del modelo final resultante, se obtuvieron estadísticos de la capacidad de predicción del modelo e intervalos de predicción de la resistencia a la fractura.

Resultados: En la validación cruzada resultó un R^2 promedio de los folds: $CVR^2 = 0,971$. El algoritmo seleccionó un modelo con entre ocho y doce funciones base y segundo grado de interacción. Las gráficas de residuales no indican anomalías en el modelo. La restricción a la deformación plástica (α) resultó ser el predictor más importante. En efecto, los nodos de las funciones base resultaron $\alpha = 10$, $\alpha = 31$ y $\alpha = 83$. El segundo predictor en importancia fue dP_{max} y en tercer lugar el parámetro de comportamiento plástico s_y . En relación a la predicción, el coeficiente R^2 sobre los conjuntos de evaluación (validación cruzada) registró valor de 0,967 y sobre el conjunto de test 0,984.

Conclusiones: La metodología utilizada mostró una buena capacidad para modelizar el fenómeno de fractura. Los nodos del predictor α pueden ser interpretados en términos de los mecanismos actuantes en el proceso de fractura. Los modelos MARS mostraron capacidad de predicción de resultados consistentes con valores obtenidos experimentalmente para este material.

Palabras Clave: Regresión no paramétrica, Modelos MARS, Fractura, Polipropileno.

¹e-mail:luis.arenas@faea.uncoma.edu.ar

Detección de fibrilación auricular en electrocardiogramas de una derivación y de corta duración

Marco Scavino ¹, Virginia Estragó ², Matías Muñoz ³, Andrés Castrillejo ⁴, Ramón Álvarez-Vaz ⁵

^{1,5} Instituto de Estadística, FCEA - Universidad de la República

^{2,3} Comisión Honoraria para la Salud Cardiovascular (CHSCV), Presidencia de la República, Uruguay

⁴ Proyecto ANII Fondo Sectorial de Investigación a partir de Datos 1_2018_1_154651

En este trabajo desarrollamos los procedimientos basados en algoritmos de aprendizaje automático diseñados para la identificación de una patología cardíaca, la fibrilación auricular (FA), a partir de la señal electrocardiográfica (ECG) de una derivación y de corta duración registrada con un dispositivo de tecnología electrónica móvil (DTEM).

El dispositivo móvil tiene un método interno de clasificación que permite un análisis instantáneo del registro de ECG en ritmo normal, posible FA, o sin clasificar. Visto la importancia del diagnóstico temprano de patologías cardiovasculares tales como la FA, nuestro objetivo es mejorar la capacidad de clasificación del DTEM el cual, en la práctica, deja sin clasificar un conjunto relevante de registros de ECG.

A tal efecto, utilizamos métodos de extracción de características, basados en técnicas de filtrado y de análisis de la morfología de la señal ECG, detectando, en particular, el llamado *complejo QRS* de cada latido del corazón y ubicando el respectivo *pico R*, siendo la medida del intervalo RR, que transcurre desde el inicio de una onda R hasta el inicio de la onda siguiente, un aspecto clínicamente relevante [1].

Analizamos el desempeño predictivo de técnicas de clasificación supervisada, tales como árboles generados por partición recursiva y bosques aleatorios, empleadas en combinación con los métodos de extracción de características de la señal ECG, a través del entorno de programación **R**. Consideramos un banco de entrenamiento internacional, constituido para la competencia 2017 “PhysioNet/Computing in Cardiology (CinC)” (<https://physionet.org/content/challenge-2017/1.0.0/>) [2], y un conjunto de prueba de registros de ECG de una población de adultos mayores de Uruguay, generado en 2019 en virtud de una colaboración entre la CHSCV y el Plan Ibirapitá.

Los resultados obtenidos se interpretan a través de los diagnósticos disponibles de los ECG realizados por expertos cardiólogos clínicos.

[1] Kannathal N., Acharya U.R., Joseph K.P., Min L.C., Suri J.S. (2007) Analysis of Electrocardiograms. In: Acharya U.R., Suri J.S., Spaan J.A.E., Krishnan S.M. (eds) Advances in Cardiac Signal Processing. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-36675-1_2

[2] Clifford Gari D. et al. (2017). AF Classification from a Short Single Lead ECG Recording: the PhysioNet/Computing in Cardiology Challenge 2017. Computing in Cardiology. Vol. 44. <https://www.cinc.org/archives/2017/pdf/065-469.pdf>

Palabras Clave: Machine Learning – Inteligencia Artificial, electrocardiograma, fibrilación auricular, clasificación supervisada.

¹ marco.scavino@fcea.edu.uy

² vestrageo@cardiosalud.org

³ mmunoz@cardiosalud.org

⁴ castrillejoandres@gmail.com

⁵ ramon@iesta.edu.uy

Modelos de Machine Learning para evaluar el riesgo crediticio de personas naturales en una institución financiera de Chiclayo-Perú.

Alfonso Tesen Arroyo ¹

Departamento de Estadística - Universidad Nacional Pedro Ruiz Gallo, Chiclayo

La presente investigación tuvo como objetivo determinar que los modelos de aprendizaje de máquina evalúan eficazmente el riesgo crediticio de personas naturales de una institución financiera de Chiclayo que el modelo clásico de credit scoring estimado mediante la Regresión Logística.

La investigación es de tipo descriptivo, explicativo y predictivo, para lo cual se trabajó con la metodología CRISP- DM.

Para el desarrollo de la investigación se utilizaron los modelos de aprendizaje de máquina tales como, Árboles de Clasificación, Redes Neuronales, Maquinas de Soporte Vectorial y el modelo clásico de la Regresión Logística; la base de datos estuvo constituida por 2464 clientes, de los cuales se utilizó el 70 % de la base para el entrenamiento y el 30 % restante para la validación. Para la comparación de los modelos, se utilizó la Matriz de Confusión y la curva ROC, determinando que el mejor modelo de clasificación global en la etapa de entrenamiento fue las Redes Neuronales con un 81.10 % y 82 % en la etapa de validación; mientras que el mejor modelo de estimación del riesgo crediticio se obtuvo mediante el árbol de decisión para nuestros datos planteados con un 35.30 % y 32.21 % en las etapas de entrenamiento y validación respectivamente.

Finalmente se concluyó que los modelos de aprendizaje de maquina evalúan mejor el riesgo crediticio que el modelo de enfoque paramétrico, como es el de la Regresión Logística para nuestros datos financieros.

Palabras Clave: Máquinas de Aprendizaje, Machine Learning, Regresión Logístico, Redes Neuronales Artificiales, Máquinas de Soporte Vectorial

¹e-mail: atesen@unprg.edu.pe

Tema 13: Métodos Estadísticos

Análisis de Datos Simbólicos: Una nueva forma de pensar en Data Science

Mallea, Lilian Adriana¹

Departamento de Matemática, Universidad Nacional de San Juan

La ciencia de datos, considerada como una ciencia en sí misma, es en términos generales, la extracción de conocimiento de los datos. Data Mining es una poderosa tecnología con gran potencial para extraer tal conocimiento. Sin embargo, desde el punto de vista estadístico, sus herramientas sólo han sido desarrolladas para trabajar con matrices de datos clásicas, es decir, donde cada unidad es individual y las variables toman un único valor para cada individuo. El análisis de datos simbólicos (ADS) brinda una nueva forma de pensar en Data Science al extender la entrada estándar a un conjunto de clases de entidades individuales. Por lo tanto, las clases de una población dada se consideran unidades de una población de nivel superior a estudiar. Tales clases a menudo representan las unidades reales de interés. Para tener en cuenta la variabilidad entre los miembros de cada clase, las clases se describen por intervalos, distribuciones, conjunto de categorías o números que a veces se ponderan. De esa manera, obtenemos nuevos tipos de datos, llamados "simbólicos", ya que no se pueden reducir a números sin perder información sobre la variabilidad interna. El primer paso en ADS es construir la tabla de datos simbólicos donde las filas son clases y las variables pueden tomar valores simbólicos. El segundo paso es estudiar y extraer nuevos conocimientos de estos nuevos tipos de datos mediante al menos una extensión de Estadística Computacional y Data Mining a datos simbólicos. ADS es un nuevo paradigma que abre un vasto dominio de investigación y aplicaciones al proporcionar resultados complementarios a los métodos clásicos aplicados a los datos estándar. Brinda respuestas a grandes volúmenes de datos (big data) y datos complejos, ya que los primeros se pueden reducir y resumir por clases y los datos complejos, con múltiples tablas de datos no estructurados y variables no apareadas, se pueden transformar en una tabla de datos estructurada con variables apareadas de valores simbólicos. A lo largo de las últimas tres décadas se han extendido distintos métodos del análisis clásico de datos al simbólico; entre los que se pueden citar: análisis de componentes principales, Clustering, análisis discriminante, regresión, métodos paramétricos y series temporales. El propósito de esta comunicación es mostrar el potencial de los métodos de ADS e ilustrar algunos con datos de COVID 19 publicados en el sitio <https://github.com/owid/covid-19-data>. En primer lugar se han filtrado los datos correspondientes a los países de América. Luego, para construir la tabla simbólica se ha realizado una agrupación temporal de datos, tomando como objeto simbólico el país y seleccionado sólo las variables referentes al total de casos, total de muertes, nuevos casos, nuevas muertes y sus relativos por millón de habitantes. También se consideraron otros indicadores referentes a la medidas de restricción de cada país frente a la pandemia, y proporción de habitantes con algunas comorbilidades. Se ha llevado a cabo una descripción simbólica de países de Latinoamérica y una clasificación no supervisada utilizando el Método de Clustering Dinámico (Verde, De Carvalho and Lechevallier, 2001) Este método determina iterativamente una serie de particiones que se mejora en cada paso de acuerdo a un criterio de agrupamiento. El algoritmo se basa en:

- prototipos para representar los clusters,
- funciones de proximidad para asignar los elementos (objetos simbólicos) a los clusters en cada etapa.

La clasificación muestra que los países se agrupan de acuerdo al total de casos y nuevos casos de Covid19 y sus correspondientes relativos por millón de habitantes. Los resultados obtenidos quizás no serán relevantes en un futuro, dada la situación dinámica de la pandemia, sin embargo lo que se pretende es mostrar la ventaja de estos métodos, al permitir tener en cuenta la variabilidad propia de los datos.

Palabras Clave: Análisis, Datos, Simbólicos, Clustering Simbólico

¹e-mail: lamallea@gmail.com, lamallea@ffha.unsj.edu.ar

Una función de covarianza de Matérn generalizada para modelar procesos espaciales bivariados

Alfredo Alegría¹, Xavier Emery², Emilio Porcu³

¹*Departamento de Matemática, Universidad Técnica Federico Santa María, Chile*

²*Departamento de Ingeniería en Minas, Universidad de Chile, Chile;
Advanced Mining Technology Center, Universidad de Chile, Chile*

³*Department of Mathematics, Khalifa University, United Arab Emirates*

Cuando se analizan datos espaciales bivariados, es fundamental modelar apropiadamente las funciones de covarianza marginales y cruzadas. En algunas ocasiones, las covarianzas cruzadas experimentales pueden exhibir características complejas. Por ejemplo, su valor máximo no necesariamente se alcanza en el origen (lo que se conoce como el efecto *cross-dimple*). Como los modelos que existen actualmente en la literatura no permiten capturar este tipo de características, en este trabajo se propone una generalización del clásico modelo de covarianza de Matérn con el objetivo de obtener una familia de covarianzas que admite un efecto cross-dimple. El modelo propuesto está basado en una modificación de la representación espectral del modelo de Matérn tradicional. Los resultados obtenidos son aplicados en datos de minería y geoquímica.

Palabras Clave: Cross-Dimple; Densidad Espectral; Distribución Gamma Inversa; Modelo de Matérn; Procesos Espaciales Bivariados.

¹e-mail: alfredo.alegria@usm.cl

²e-mail: xemery@ing.uchile.cl

³e-mail: emilio.porcu@ku.ac.ae

Agreement assessment between two measurement systems using robust P-splines

Felipe Osorio, Manuel Galea¹

Departamento de Matemática, Universidad Técnica Federico Santa María, Chile.

Departamento de Estadística, Pontificia Universidad Católica de Chile.

Assessing agreement between two measurement instruments which varies in pricing, fastness and other features, has been of growing interest in several areas of knowledge. Some methods assume that the differences of the paired measurements come from a normal distribution. In this work, we develop the probability of agreement to compare measurement systems using a semiparametric approach. Because the assumption of normality of the differences of the paired measurements, is unrealistic in many situations, we focus on the use of penalized splines considering the scale mixture of normal distributions to model non-normal measurements. This class of models is flexible and allows to model datasets with high levels kurtosis, and potential outliers.

Palabras Clave: Measurement systems comparison, Probability of agreement, Robust P-splines.

¹e-mail: felipe.osorios@usm.cl, mgalea@mat.uc.cl

Vectores de capacidad de procesos multivariados. Sobre sus propiedades distribucionales

Dra. Daniela F. Dianda, Dra. Marta B. Quaglino¹

Instituto de Investigaciones Teóricas y Aplicadas, Escuela de Estadística, Universidad Nacional de Rosario

La importancia del análisis de capacidad de procesos deriva de la filosofía de adecuación a los requerimientos del cliente, presente en todas las concepciones de calidad. Satisfacer los requerimientos del mercado es una premisa fundamental para toda empresa y para ello es necesario que el comportamiento de las características relevantes de los productos o servicios ofrecidos, sean monitoreadas constantemente y que sus niveles de operación habitual sean contrastados contra los requisitos que sobre ellas se impongan. En la actualidad, es habitual que las especificaciones se establezcan sobre diversas características de calidad de un mismo producto, de modo que su conformidad queda determinada por el comportamiento conjunto de todas ellas. Es así que la capacidad del proceso de satisfacer los requerimientos debe analizarse desde una perspectiva multivariada, enfoque que se encuentra aún en pleno desarrollo. En este trabajo se aborda uno de los enfoques que propone la literatura para analizar la capacidad de un proceso multivariado, específicamente en la línea que propone la creación de medidas multidimensionales, denominadas *vectores de capacidad*.

La idea de crear un indicador multidimensional surge como respuesta a la dificultad que presentan las medidas unidimensionales para resumir en un único valor todos los aspectos involucrados en un proceso multivariado y alertar al usuario sobre la causa de la falta de capacidad, en caso de ocurrir. Los vectores de capacidad fueron diseñados con la idea de que cada componente relate sobre un aspecto del proceso en particular.

La mayoría de las propuestas existentes en esta línea, consiste en vectores de tres componentes. La primera evalúa el desempeño del proceso en variabilidad, cuantificado a través del cociente entre el volumen de la región de especificación y el volumen de la región de variación natural del proceso. La segunda componente evalúa el centrado del proceso y corresponde a la probabilidad asociada de un test para la media de un vector aleatorio. La tercera componente compara la localización de la región de variación natural del proceso con la región de tolerancia.

En este trabajo se presentan los resultados de estudios preliminares sobre el comportamiento muestral de tres de los vectores de capacidad más divulgados, para el caso de especificaciones de calidad bilaterales: el vector *VCH* de Hubele *et al.* (1991), el vector de *MCPV* de Shahriari *et al.* (1995) y el vector *NMPCV* de Shahriari y Abdollahzadeh (2009).

Los resultados hallados permiten derivar recomendaciones de uso de cada alternativa, aspecto también en carencia en la literatura actual.

Palabras Clave: Control de calidad, Análisis de capacidad multivariado, Distribuciones muestrales.

¹e-mail: ddianda@fcecon.unr.edu.ar, mquaglino@fcecon.unr.edu.ar

Tema 14: Misceláneos

Efecto de las restricciones en los empates en el Test de Friedman

Carlos López-Vázquez¹, Andrómaca Tasistro², Esther Hochsztain³ ¹

¹Laboratorio LatinGEO IGM+ORT, Universidad ORT Uruguay, ²Investigador independiente, ³Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay

El Test de Friedman [1937] provee medios para dilucidar si en un concurso de vinos con N jueces y k vinos hay evidencia estadística que rechace la hipótesis nula de que todos los vinos son equivalentes. Para ello se calcula primero el valor del estadístico χ_r^2 . Si k y N son moderados o grandes Friedman ofrecía aproximaciones asintóticas. Para k y N pequeños no valen las mismas, por lo que se ofrecían tablas. En su formulación original los jueces no podían asignar empates. A partir de Conover [1980], se comenzaron a utilizar formas alternativas para el estadístico χ_r^2 que consideraban la posibilidad de empates, manteniendo las mismas aproximaciones asintóticas. Igual que antes, esas aproximaciones no eran correctas para k y N pequeños. Sólo recientemente López-Vázquez, Tasistro y Hochsztain [2021] presentaron las primeras tablas exactas para el caso con empates, válidas para un conjunto reducido de valores de k y N . Hoy es posible utilizar una única expresión general del estadístico χ_r^2 válida para casos con y sin empates, y recurrir a las tablas adecuadas según corresponda al caso. En principio podría pensarse que eso constituye el punto final al problema, en la medida que se trata de tablas exactas.

Ello no es así ya que el tipo de empates no ha sido considerado hasta ahora. Si existiera algún problema en que no todos los posibles empates (de a 2, de a 3, ..., de a k) o alguna combinación de ellos no sean admisibles, entonces cabe preguntarse si pueden aún utilizarse las tablas que ignoran ese detalle. En este trabajo se demuestra que, en el caso en que el problema impida que ciertos empates puedan darse, las tablas generales no coinciden con las que se deberían utilizar. La demostración es de tipo constructivo: se describe el proceso seguido para la construcción de las tablas del estadístico imponiendo restricciones arbitrarias a los empates, y se confirma que no se obtiene el mismo resultado. Se han realizado los cálculos exactos en unos pocos casos, por lo que no pueden establecerse reglas generales para cuantificar las diferencias que ocurren frente a la tabla con empates sin restricciones. Para cada restricción que se imagine deberá construirse la tabla específica. En la medida que las posibles restricciones aumentan con k no parece viable construirlas a priori.

Para ilustrar las consecuencias se discute en este trabajo un ejemplo basado en los criterios habituales para empates de una competición olímpica. Si se restringe a tres el número de medallas o diplomas a entregar, las tablas requeridas para rechazar la hipótesis nula de que hay no diferencias entre los competidores no pueden ser las estándar.

Referencias

- Conover, W. J., 1980. Practical Nonparametric Statistics 2nd. Edition, Wiley.
- Friedman, M., 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. Journal of the American Statistical Association, 32, 200:675-701
- López-Vázquez, C., Tasistro, A. and Hochsztain, E., 2021 Exact tables for the Friedman rank test: Case with ties. Chilean Journal of Statistics, 12, 1:23-35

Palabras Clave: Estadística no paramétrica, Friedman Test, Empates

¹e-mail: carloslopez@uni.ort.edu.uy, tasistro@gmail.com, estherhochsztain@gmail.com

Una propuesta de estimación en el Modelo Nulo de dos niveles

Piergentili, M. V.^[1], Schlaps, E.^[2], Martín, M. C.^{[1] [3]}¹

^[1] *Departamento de Matemática, Universidad Nacional del Sur*

^[2] *Instituto de Ciencias Polares, Ambiente y Recursos Naturales, Universidad Nacional de Tierra del Fuego, Antártida e Islas del Atlántico Sur*

^[3] *Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa*

Los modelos que incluyen información jerárquica en su definición son conocidos como **Modelos Multinivel**. En estos modelos se identifican tanto parámetros que representan la parte fija como la parte aleatoria del mismo. Los parámetros de la parte fija, o previsible, son las pendientes y el intercepto que caracterizan los efectos medios de la población; mientras que los parámetros asociados a la parte aleatoria corresponden a las varianzas y covarianzas de todos los niveles, llamadas comúnmente *componentes de varianza*.

Se considera el **Modelo Aditivo de dos Niveles** $y_{ij} = \beta_{0j} + e_{ij}$, $i = 1, \dots, n_j, j = 1, \dots, m$, denominado *Modelo Nulo* pues no incluye variables predictoras en su definición, es decir, sólo posee efectos aleatorios en ambos niveles; el subíndice i hace referencia a las unidades de nivel 1 y el subíndice j a las unidades de nivel 2, permitiendo descomponer la varianza en dos componentes independientes: σ_e^2 que es la varianza del residuo del primer nivel y $\sigma_{u_0}^2$ que es la varianza del residuo del segundo nivel.

En este modelo de 2 niveles, β_{0j} se considera una variable aleatoria en el segundo nivel, y se tiene que:

- $\beta_{0j} = \beta_0 + u_{0j}$, con $u_{0j} \sim N(0, \sigma_{u_0}^2)$ (o sea, el intercepto β_{0j} se descompone en una parte fija, β_0 que representa la media general de la población y una parte aleatoria u_{0j});
- β_{0j} representa las v.a.i.i.d. del segundo nivel, i.e., $\beta_{0j} \sim N(\beta_0, \sigma_{u_0}^2)$ y $E(u_{0j_1} u_{0j_2}) = 0, \forall j_1 \neq j_2$.

Además:

- e_{ij} representa las v.a.i.i.d. del primer nivel, i.e., $e_{ij} \sim N(0, \sigma_e^2)$ y $E(e_{i_1 j} e_{i_2 j}) = 0, \forall i_1 \neq i_2$;
- Existe independencia entre los errores del primer y segundo nivel, i.e., $E(e_{ij} u_{0j}) = 0, \forall i, j$.

El trabajo presenta un estudio teórico detallado de los estimadores, y sus propiedades, para los parámetros de la parte fija obtenidos por Mínimos Cuadrados Generalizados y una propuesta para los estimadores de las componentes de la varianza a partir del Análisis de la Varianza de un factor. Utilizando estas estimaciones se definen los residuos que intervienen en ambos niveles y que permitirán evaluar los supuestos del modelo.

Palabras Clave: modelo nulo multinivel, componentes de varianza, métodos de estimación.

¹e-mail: virginia.piergentili@uns.edu.ar, eschlaps@untfd.edu.ar, maritamartin11@gmail.com

Origen de los autores y temas tratados en los artículos recientes de las revistas latinoamericanas en estadística de la corriente principal de la ciencia

Rafael E. Borges 1, Jesús W. Salinas 2, Lida R. Fonseca 3¹

1 Escuela de Estadística, Universidad de Los Andes, Mérida, Venezuela, 2 Departamento de Estadística e Informática, Universidad Nacional Agraria La Molina, Lima, Perú, 3 Facultad de Ingeniería, Fundación Universitaria Los Libertadores, Bogotá, Colombia

En este trabajo se analizan los artículos incluidos en las revistas científicas en el área de estadística incluidas en los índices de la corriente principal de la ciencia (Scopus y Web of Science) para el período comprendido entre los años 2019 y 2021. Las revistas consideradas en el análisis fueron: *Latin American Journal of Probability and Mathematical Statistics* (ALEA), encontrándose en el período se publican 3 volúmenes de cada revista, publicándose de ésta, 6 números y 145 artículos, *Brazilian Journal of Probability and Statistics* (BJPS), con 11 números y 124 artículos, *Chilean Journal of Statistics* (ChJS), con 5 números y 27 artículos, la *Revista Colombiana de Estadística* (RCE), con 6 números y 44 artículos, y la *Revista Investigación Operacional* (RIO), con 15 números y 184 artículos, con lo cual se considera un total de 524 artículos.

De estos artículos se analizaron los autores de los mismos, llevándose a cabo un análisis de minería de texto a nivel exploratorio para identificar los países de adscripción de los mismos, según la importancia relativa en cada una de las revistas y se analizan las redes de cooperación observadas en las mismas, y se analizaron las palabras clave en inglés (keywords) los resúmenes en inglés (abstracts) usando un análisis de minería de texto a nivel exploratorio para identificar la temática de los trabajos que se están publicando en estas prestigiosas revistas.

Al analizar las nubes de palabra para los autores, se puede observar que existen diferencias en la nacionalidad de los autores para cada revista, observándose, que la nacionalidad de los autores que más publican en la BJPS y la ChJS es la brasilera, en ALEA publican principalmente franceses, alemanes, estado unidenses y chinos, en la RCE publican principalmente colombianos y brasileros, mientras que el la RIO están publicando principalmente ecuatorianos, hindúes, cubanos y mexicanos.



Y al analizar las nubes de palabra para las palabras claves, se observa que ALEA tiene una tendencia hacia los procesos estocásticos y la probabilidad, la RIO tiene una tendencia hacia la computación y las tres revistas restantes tratan temáticas más relacionadas con la estadística

Palabras Clave: Revistas estadísticas. Minería de texto, Nubes de palabras.

¹e-mail: borgesr@gmail.com, jsalinas@lamolina.edu.pe, lrfonseca@libertadores.edu.co

APLICACIÓN DE INGENIERIA ESTADISTICA A UN ESTUDIO DE SALUD AUDITIVA

Andrea F. Righetti, Silvia Joeques, Sergio M. Buzzi, Rodrigo Porta, María Virginia Bederian, Ester C. Biasoni, Mario R. Serra.¹

Instituto de Estadística y Demografía, Facultad de Ciencias Económicas, Universidad Nacional de Córdoba

Centro de Investigación y Transferencia en Acústica (CINTRA) Unidad Asociada de CONICET, Facultad Regional Córdoba, Universidad Tecnológica Nacional

La Asociación Internacional de Ingeniería Estadística (ISEA por sus siglas en inglés), adoptó recientemente la definición de ingeniería estadística como; “una disciplina dedicada al arte y la ciencia de resolver problemas complejos que requieren datos y análisis de datos”. Esta disciplina emergente se focaliza en el estudio de la integración sistemática de conceptos, métodos y herramientas estadísticas, a menudo con otras disciplinas relevantes, para resolver problemas complejos y muchas veces no estructurados. Con la finalidad de construir un enfoque unificado, estadísticos, ingenieros y académicos trabajan de manera colaborativa y proponen diferentes fases para llevar a cabo estos proyectos. En el presente trabajo se describe un caso de aplicación de ingeniería estadística a la resolución de un problema de salud auditiva. Específicamente debía encontrarse una respuesta científica al alto porcentaje de jóvenes rechazados en el ingreso laboral en Argentina por problemas auditivos sin justificación clínica. Este problema dio origen a un Programa multidisciplinario y multiinstitucional que necesitó ser abordado desde diferentes áreas, incluyendo aspectos audiológicos, psicosociales, acústicos, genéticos y estadísticos. El análisis de los datos relevados permitió establecer una estrategia educativa de intervención para concientizar a los jóvenes sobre los efectos nocivos del ruido para la audición.

Palabras Clave: control de calidad. ingeniería estadística. salud auditiva. hipoacusia. bioestadística.

¹e-mail: arighetti@unc.edu.ar

Bootstrapping Regression Models with Locally Stationary Disturbances

Guillermo Ferreira ¹, Jorge Mateu , Jose A. Vilar , Joel Muñoz

Department of Statistics, Universidad de Concepción

A linear regression model with errors following a time-varying process is considered. In this class of models the smoothness condition both in the trend function and in the correlation structure of the error term ensures that these models can be locally approximated by stationary processes, leading to a general class of linear regression models with locally stationary errors. We focus here on the bootstrap approximation to the distribution of the least squares estimator (LSE) for such class of regression models. We compare and discuss the results on both the classical and bootstrap confidence intervals through an intensive simulation study. The trend is also discussed through a real data analysis on time series of monthly inflation in US with locally stationary errors.

Palabras Clave: Bootstrapping, Local stationarity, Non-stationarity, Time-varying models.

¹e-mail: gferreir@udec.cl

Uso de información auxiliar en gráficos de control para la media en fase II. Análisis comparativo.

Daniela F. Dianda, Vera Teves, Marta B. Quaglino¹

Instituto de Investigaciones Teóricas y Aplicadas, Escuela de Estadística, Universidad Nacional de Rosario

Los modelos de gestión empresarial tienen como foco principal la calidad, la satisfacción del cliente y la mejora continua de sus procesos. El logro de estas aspiraciones está directamente relacionado al empleo de herramientas que permitan analizar objetivamente la información de todos los procesos involucrados para la toma de decisiones. La variabilidad a la que están expuestos los procesos pueden alterar la calidad del producto final, por lo que es necesario tomar acciones que garanticen la estabilidad de los mismos. Los diagramas o gráficos de control están específicamente diseñados para identificar la presencia de fuentes de variabilidad que ocasionan inestabilidad en el funcionamiento del proceso.

La implementación de los gráficos de control en un proceso real, consta de dos fases bien diferenciadas: la fase I o fase retrospectiva, cuyo objetivo es establecer los niveles de operación naturales del proceso, y la fase II o prospectiva, dedicada al monitoreo en línea del mismo. En general, la implementación de los gráficos de fase II asume que los parámetros de la distribución subyacente son conocidos. Esta suposición simplifica el desarrollo y evaluación de los diagramas de control, aunque en la práctica estos parámetros suelen ser estimados con los resultados de la fase I. Cuando se usan estimaciones en lugar de los parámetros conocidos, la variabilidad de los estimadores puede afectar el desempeño del gráfico, respecto del esperado para los gráficos diseñados con parámetros conocidos. En particular, para los gráficos Shewhart, la utilización de estimaciones de los parámetros genera una disminución en la potencia del gráfico en la fase II, lo que ocasiona que sean menos sensibles a cambios pequeños o moderados del proceso.

La utilización de información de variables auxiliares es una estrategia que ha sido utilizada para mejorar el desempeño de los gráficos de control y ha permitido diseñar adaptaciones de los gráficos Shewhart tradicionales, mejorando su rendimiento y aprovechando sus ventajas. Una de tales adaptaciones es la propuesta de los autores Shabbir y Awan (2015), quienes incorporan el estimador *difference in difference* en el diseño de un gráfico tipo Shewhart, con mejor sensibilidad para detectar cambios pequeños y moderados en la media del proceso.

En este trabajo se compara el desempeño del gráfico resultante, frente a diversas magnitudes de cambio en la media del proceso, estimando los parámetros necesarios a partir de muestras de los tamaños usualmente recomendados, con el de los gráficos Shewhart tradicionales y con las estrategias de gráficos ponderados en el tiempo, generalmente recomendados para el monitoreo de los procesos en fase II.

Palabras Clave: Control de procesos, Estimación con variables auxiliares, Fase I y Fase II.

¹e-mail: ddianda@fcecon.unr.edu.ar, veranicoleteves@gmail.com, mquaglino@fcecon.unr.edu.ar

Tema 15: Muestreo

Análisis de los textos publicados en Facebook por los concejales de Rosario, Argentina, durante el año 2019

Luisina Rubio, Julián Crucella, Marcos Prunello¹

Escuela de Estadística, Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario

En el último tiempo las redes sociales han alterado las interacciones humanas permitiendo a las personas tanto compartir sus opiniones y experiencias, como reaccionar sobre lo publicado por otros. La popularidad alcanzada por las redes sociales se ha trasladado rápidamente al terreno de la política, donde la mayoría de los partidos, gobernantes y/o funcionarios cuentan con alguna de estas herramientas para conectarse con la ciudadanía.

En este trabajo se analizó la comunicación de los concejales de Rosario en el año electoral 2019, entre los meses de febrero y noviembre. Se tomaron los mensajes incluidos en todas las publicaciones realizadas en sus páginas públicas de Facebook y, luego de un proceso de depuración, se emplearon técnicas para el análisis de texto, como nubes de palabras y grafos de correlaciones. Mediante estas herramientas fue posible caracterizar el contenido de las comunicaciones de cada bloque político, identificando aspectos locales relacionados a las campañas y también vinculaciones con asuntos o actores políticos, ya sean estos nacionales o regionales.

Finalmente, se empleó la modelización de tópicos con asignación latente de Dirichlet (*LDA topic modeling*), siendo posible señalar tres núcleos temáticos en los mensajes publicados: jerga de campaña, propuestas políticas y cuestiones de género y derechos humanos. De esta forma, se pudo identificar la preponderancia de estos temas en el discurso de cada bloque de concejales.

Palabras Clave: Facebook, elecciones, concejo municipal, análisis de texto, topic modeling, asignación latente de Dirichlet

¹e-mail: luisinarubio@gmail.com, info@inmediata.com.ar, mprunello@fcecon.unr.edu.ar

Aumentando el tamaño muestral mediante el uso de doubles ventanas móviles

Virgilio Pérez, Cristina Aybar y Jose M. Pavía¹

GIPEyOP, Universidad de Valencia

Cuando se realiza una investigación, a menudo es imposible, o demasiado caro, recoger datos de todas las personas o elementos que nos interesan. En su lugar, tomamos una muestra de la población de interés y predecimos cómo se comporta la población a partir de esa muestra.

Hay muchos aspectos que influyen en lo mucho (o poco) que una muestra se parece a la población y, por tanto, en la validez y fiabilidad de nuestras conclusiones. Uno de los conceptos clave que hay que tener en cuenta al realizar una investigación es el tamaño de la muestra. Es bien sabido que a medida que se dispone de más información (mayor tamaño de la muestra), se reduce la incertidumbre. Por ello, en este estudio nos hemos centrado en cómo ampliar el tamaño de la muestra sin necesidad de obtener nuevos datos, y sin apenas aumentar los niveles de error asociados.

Como decimos, a veces no podemos validar nuestras investigaciones porque tenemos pocos datos (tamaños de muestra pequeños), una circunstancia que provoca mucha volatilidad. Para solucionar este problema, proponemos utilizar información adyacente, es decir, admitir que los sujetos cercanos se comportan de forma similar. Para verificar este supuesto, hemos utilizado una gran base de datos con más de 150 variables y más de 700.000 observaciones, a lo largo de 30 años. Hemos comprobado que, efectivamente, hay ciertas variables temporales que evolucionan de forma muy suave, como la ideología política, entre otras, por lo que no es descabellado establecer clusters de datos (o pools de datos). El método propuesto describe el uso de las ventanas móviles, logrando el objetivo descrito de aumentar el tamaño de la muestra sin apenas aumentar el error de estimación.

Palabras Clave: ventanas móviles, tamaño muestral, error de estimación.

¹e-mail: virgilio.perez@uv.es, cristina.aybar@uv.es, pavia@uv.es

Impacto de la anonimización en datos de estadísticas oficiales: Un caso de aplicación a un indicador de referencia

Bussi, Javier ; Marí, Gonzalo ; Méndez, Fernanda¹

*Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística, Facultad de Ciencias
Económicas y Estadística, Universidad Nacional de Rosario*

Los operativos estadísticos en el ámbito de las estadísticas oficiales tienen por objetivo recolectar datos de unidades de observación que pueden estar constituidas por distintas entidades, tales como personas, hogares, empresas u otros objetos. Estos datos proveen información relevante que contribuye a la caracterización del estado de situación de las políticas implementadas y también a la planificación de políticas futuras basadas en la toma de decisiones inferidas a partir de dicha información. Este es uno de los objetivos principales de las estadísticas oficiales. Es evidente el crecimiento en los últimos años de una marcada tendencia a que estos datos estén disponibles, de manera accesible y rápida, al conjunto de la sociedad, de manera tal que sus distintos actores puedan realizar sus propios análisis y obtener sus propias conclusiones, pudiendo aportar una óptica particular de los distintos aspectos estudiados. Por lo tanto, en años recientes, el concepto de datos abiertos se ha extendido y generalizado de manera tal, que existe un gran número de portales oficiales que contienen datos gratuitos, reutilizables y sin restricciones de derecho de autor, a los cuales el público en general puede acceder fácilmente. A su vez, en la Argentina existe el marco legal constituido principalmente por la Ley N° 17.622 que menciona la obligación de respetar el secreto estadístico en la divulgación de los datos provenientes de los operativos que se llevan a cabo en el ámbito de las estadísticas oficiales en cuanto a preservar el anonimato de las unidades de información. Es evidente entonces que existe un conflicto entre ambas condiciones, por un lado los datos abiertos y por el otro, el secreto estadístico. Esta contraposición representa un desafío importante a tener en cuenta para buscar soluciones que permitan brindar información que debería ser pública sin violar el secreto estadístico de las entidades involucradas en suministrar datos. Surge entonces la necesidad de desarrollar alternativas que permitan satisfacer ambos objetivos y es por este motivo, que se estudian distintas variantes que se engloban dentro de los llamados métodos de anonimización. Estos métodos permiten, en cierto grado, proteger la identidad de las unidades, es decir, que no pierdan el anonimato, de manera tal que la información pueda estar disponible con un riesgo bajo de divulgación. Estos métodos pueden modificar los datos originales (perturbadores) o no (métodos no perturbadores). En este trabajo se presentan métodos de anonimización con el fin de evitar que se pueda detectar, utilizando información secundaria, la identidad de las unidades, lo que generaría un incumplimiento de la ley de secreto estadístico por parte del organismo que publica la información. Se evalúan las técnicas de perturbación que adicionan un ruido aleatorio a los datos (correlacionado o no) y la de microagregación, que asigna un valor representativo a todas las unidades que forman parte de un grupo de unidades determinado por la cercanía de las mismas. Se comparan los métodos propuestos a través de medidas que muestran la modificación que producen en las estimaciones de ciertos parámetros relacionados con un indicador de referencia, en este caso el Coeficiente de Gini. En el presente trabajo, se evalúan ciertos escenarios de anonimización sobre la base usuario de la Encuesta Nacional de Gastos de Hogares (ENGHO) publicada por el INDEC, utilizando las bases de pesos replicados para la estimación de ciertos parámetros que cuantifican el error muestral. En general, los métodos que adicionan un ruido aleatorio no correlacionado presentan mejores resultados que sus competidores, y más consistentes que los que produce el método de microagregación. Este método parece estabilizarse y lograr resultados más consistentes en los escenarios planteados al incrementarse el tamaño muestral.

Palabras Clave: Estadísticas Oficiales, Anonimización, Indicador de referencia

¹e-mail: jbussi@fcecon.unr.edu.ar

Impacto de información no pública en el cálculo de varianzas en la Encuesta Continua de Hogares

Ana Coimbra, Mauricio Pittamiglio, Juan José Goyeneche¹

Instituto de Estadística, FCEA, Universidad de la República, Uruguay

En el marco del convenio entre la Facultad de Ciencias Económicas y de Administración (FCEA) y el Observatorio Territorio Uruguay (OTU) de la Oficina de Planeamiento y Presupuesto (OPP), se estudió el desempeño de algunos indicadores socioeconómicos anuales que la OTU presenta en su página web (<https://otu.opp.gub.uy/>).

El desempeño de estos indicadores se mide a partir del error estándar de sus estimaciones puntuales, y el objetivo de este trabajo es evaluar este desempeño en tres escenarios posibles.

Estos indicadores socioeconómicos se calculan a partir de la Encuesta Continua de Hogares (ECH) que es llevada a cabo por el Instituto Nacional de Estadística (INE) de Uruguay. Hasta el año 2019, la ECH era una encuesta mensual de tipo cross-section cuyo diseño muestral es aleatorio, estratificado y en dos etapas de selección. La población del Uruguay es particionada en estratos y en cada uno se seleccionan, en una primera etapa, conglomerados de zonas censales (Unidades Primarias de Muestreo – UPM), bajo un diseño aleatorio, sistemático y proporcional al tamaño (PPS), utilizando como medida de tamaño (MOS) la cantidad de viviendas particulares según el censo de población y viviendas del año 2011. En una segunda etapa, se seleccionan, bajo un diseño simple, cinco viviendas en cada una de las UPM seleccionadas.

Los tres escenarios se diferencian en la información disponible del diseño muestral con la que cuenta el usuario. El nivel con menos información disponible es el que utiliza las bases públicas del INE, en la que no se especifica a qué estrato ni a que UPM pertenece el individuo (hogar o persona). Un segundo nivel es cuando se conoce la UPM a la que pertenece el individuo, pero no su estrato y, por último, la opción más completa, es cuando se dispone de la UPM y el estrato al que pertenecen todos los individuos.

En este trabajo se presenta el desempeño de 5 de los indicadores que calcula la OTU según los tres escenarios posibles de información disponible.

Cabe aclarar que a partir del año 2020 el diseño de la ECH cambió radicalmente y en el futuro se deberá reestudiar la vigencia de estos resultados.

Palabras Clave: Muestreo de Poblaciones Finitas, diseños muestrales complejos, cálculo de la varianza

¹e-mail: acoimbra@iesta.edu.uy, mauripityaborb@gmail.com, jjgoye@iesta.edu.uy

Muestreo a través de redes sociales: Usina de Percepción Ciudadana

Juan José Goyeneche, Leandro González¹

Instituto de Estadística, FCEA, Universidad de la República y Usina de Percepción Ciudadana

Con el objetivo de estudiar metodológicamente encuestas realizadas a través de redes sociales se realizó un proyecto conjunto de docentes del Instituto de Estadística, la diaria (<https://ladiaria.com.uy/>) y el grupo Etcétera: la Usina de Percepción Ciudadana. El proyecto surgió a principios de 2020 con el apoyo del Programa de las Naciones Unidas para el Desarrollo.

La Usina de Percepción Ciudadana tiene además entre sus objetivos producir información para promover la discusión pública e información sobre opinión pública que pueda colaborar durante la pandemia.

El principal producto es un panel de 400 adultos, DataPanel, reclutados a través de Facebook e Instagram para formar un “marco” de donde se selecciona aleatoriamente una muestra estratificada. La muestra se pondera para reproducir las proporciones por sexo y edad según las proyecciones del Instituto Nacional de Estadística. Se realizan consultas semanales mediante un bot de WhatsApp.

Se discutirán las fortalezas y debilidades del sistema, particularmente velocidad de respuesta, insesgamiento y precisión de las mediciones y algunos de los resultados más importantes que se obtuvieron.

Palabras Clave: Muestreo de poblaciones finitas, muestras no aleatorias, estimador dependiente del modelo

¹e-mail: jjgoye@iesta.edu.uy, leandrogonzalezsosa@gmail.com

Post Estratificación Electoral en Encuestas de Opinión: Elecciones Generales de Bolivia 2020

Ronal Edwin Condori Huanca¹

I.E.T.A., F.C.P.N., Universidad Mayor de San Andrés

La importancia de las encuestas de opinión con fines electorales radica tanto en su precisión sobre los resultados oficiales, los cuales siempre serán conocidos posteriormente, y en el interés coyuntural que genera sobre la sociedad, los medios de comunicación, los políticos y la academia. En particular, el diseño de muestreo, que permite definir la selección y el proceso de estimación en base a una muestra, es un componente neurálgico en esta operación estadística y sus resultados.

El diseño de muestreo de estas encuestas, aplica mayormente la estratificación, y el uso de dos o más etapas en el contexto Boliviano a lo largo de la última década. Sin embargo, la incorporación de regiones bisagra o lo que aquí denominamos como estratificación electoral, no se la ha implementado aun. Esta misma, es una alternativa viable, para mejorar la precisión de las estimaciones de encuestas de opinión. Este tipo de estratificación, no es muy utilizado en Bolivia, y tampoco en países de la región. Sin embargo es ampliamente usado en otros países, tal es el caso de Estados Unidos, donde su sistema político de elección de autoridades, permite a poner más énfasis en regiones bisagra donde hay más atención en las brechas electorales más estrechas, denominando algunas estados como "Bisagra", y centrando sus esfuerzos estadísticos en las estimaciones para estas regiones.

Usando datos históricos de resultados oficiales de eventos electorales de 2014, 2016 y 2019 de Elecciones Generales y un Referéndum, permitieron construir indicadores de variables electorales, que se usaron para la aplicación de un análisis de componentes principales previo, seguido de un análisis cluster no jerárquico para la conformación de post estratos de regiones bisagra o de amplia ventaja. Posteriormente se incluyeron estos en el diseño, además otras variables proxy del Nivel Socio Económico (NSE), como es el nivel de educación.

Esto permitió hacer ajustes de post estratificación por el método Raking a los factores de expansión originales de la 2da encuesta de intención de voto de la iniciativa "TuVotoCuenta", que fue elaborada por la UMSA en 2020, la cual es una encuesta de tipo probabilístico, aplicando un muestreo estratificado y por etapas originalmente.

Estos ajustes permitieron mejoras en métricas sobre la precisión de encuestas de este tipo. Estas métricas que son recomendadas por "The National Council of Public Polls", que fueron planteadas desde las controvertidas encuestas para las presidenciales de 1949 en EEUU. Estas métricas, como son la Desviación Media Absoluta, que redujo de 6.9 a 4.8, el Error del Margen de 6.9 a 4.9, y otras medidas más, se obtuvieron al comparar con las estimaciones originales, las cuales se usaron como contrafactual de contraste. Permitiendo verificar la viabilidad y ventaja de este tipo de muestreo, sobre los diseños convencionales.

Palabras Clave: Post-estratificación, Bisagra, Elecciones Generales, Encuestas, Raking, muestreo

¹e-mail: ronal.c.huanca@gmail.com

Tema 16: Multivariado

Método STATICO: Una aplicación en el campo de la ecología

Zuliani R., Laffitte L., Muñiz Saavedra J., Lavalle A., Bramardi S.¹

Departamento de Estadística, Universidad Nacional del Comahue. Neuquén

Numerosos estudios ecológicos tienen como interés analizar la relación entre dos tablas de datos: una de frecuencias o abundancia de especies en diversos sitios de muestreo; y por otro lado, una tabla de información de variables ambientales medidas sobre esos mismos sitios. Cuando las dos tablas se repiten en el tiempo o en diferentes situaciones se tiene una secuencia de tablas ecológicas pareadas. El método estadístico STATICO (Simier et.al.1999) puede ser empleado para analizar esta serie de pares de tablas ecológicas. El objetivo de esta técnica es obtener la parte estable de la dinámica de las relaciones entre las especies y su entorno. En esta aplicación cada par de tablas está constituida por una de frecuencias de 47 familias de macroinvertebrados (en columnas) y otra con información de 13 variables del hábitat que los circunda (en columnas) en dos momentos diferentes: 2015 y 2016. Los 8 sitios de muestreo (en filas) de cada par son los mismos y corresponden a diferentes tipos de cuerpos de agua de la Cuenca Lácar -Huan Hum. Los resultados de la aplicación muestran una relación entre las variables ambientales y familias de macroinvertebrados bastante similar en ambos años. La parte estable de las relaciones familias de macroinvertebrados-variables ambientales consiste principalmente de: (1) sobre el primer factor una combinación con un gradiente que va desde familias más tolerantes en sitios con aguas más contaminadas a las menos tolerantes asociadas a sitios con aguas menos contaminadas; (2) sobre el segundo factor familias asociadas a sitios con aguas turbias (Fig.1.A). En general las trayectorias muestran bastante similitud entre las relaciones en los dos años de estudio. Es notorio el aumento en sólidos totales en suspensión en el 2016. Se confirma la modificación en la distribución de las familias de macroinvertebrados con Chironomidae presente en ambos años; en tanto que la familia Simuliidae presente en el 2016 cuando se observa mayor turbidez en el agua. Se muestra claramente la dinámica de las relaciones familias de macroinvertebrados-variables ambientales. En los dos años se observa una moderada estabilidad del conjunto de las familias y de las variables ambientales. Las principales diferencias se dan en el sitio PP con mayor turbidez asociada con la familia Simuliidae. Por otro lado el sitio TT siendo el más contaminado con presencia de familias más tolerantes a la contaminación.(Fig.1.B). Este trabajo pretende ser un aporte para la ecología de conservación y para la ecología de cambio global. Se prevé realizar otras formas de Statico gracias a la flexibilidad que brinda el análisis de coherencia. Todos los cálculos computacionales y gráficos se realizan en el entorno de programación y análisis de datos de R.

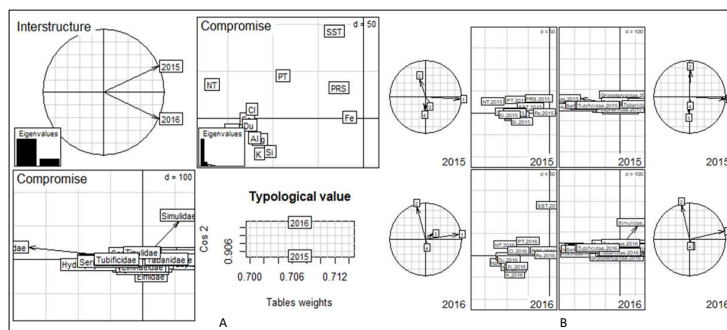


Figura 1: Interestructura-Compromiso (A). Trayectorias (B)

Palabras Clave: statis, compromiso, interestructura, análisis triádico parcial, mediciones, tablas ecológicas

Referencias:

Simier, M., L. Blanc, F. Pellegrin, and D. Nandris. (1999). *Approche simultanée de K couples de tableaux: application à l'étude des relations pathologie végétale—environnement. Revue de Statistique Appliquée in Revue de Statistique Appliquée* 47:31–46.

¹e-mail: zulianiroxanna@gmail.com, lorelaffitte@gmail.com, lavalleanrealina@gmail.com, sbramardi@mail.com

DATOS PANEL Y SU IMPUTACIÓN MÚLTIPLE. ESTUDIO CON EL ÍNDICE DE GINI

Diana Del-Callejo-Canal¹, Margarita Edith Canal-Martínez¹, Elena Vernazza² Alar Urruticoechea³, Ramón Álvarez-Vaz^{2,1}

¹*Instituto de Investigación de Estudios Superiores, Económicos y Sociales de la Universidad Veracruzana, México.*

²*Instituto de Estadística, Facultad de Ciencias Económicas y de Administración, UDELAR, Uruguay.*

³*Universidad Católica del Uruguay, Uruguay.*

INTRODUCCIÓN: Las técnicas de imputación se pueden dividir en dos: imputación simple, cuando existe un patrón monótono de datos faltantes e imputación múltiple, cuando existe un patrón arbitrario. La elección del modelo de imputación dependerá, por una parte, de las características de la tabla de datos y, por otra, de la información disponible alrededor de los datos faltantes; el patrón de los datos faltantes, el tipo de datos y de la estructura de la tabla de datos (datos panel, diseño experimental, etc.). En este trabajo, los datos utilizados son datos panel, es decir que a los individuos se les mide una o más variables a lo largo del tiempo, de tal manera que la variación en la temporalidad y la variación en los individuos resultan igual de importantes para el estudio. Considerando estas características y dado el patrón arbitrario que poseen los datos faltantes, el método de imputación que se suele utilizar es el de imputación múltiple. Este método, reemplaza cada dato faltante por un conjunto de datos aceptables (verosímiles) que representan la incertidumbre alrededor del valor real (desconocido), después de ser analizados se completa el dato faltante original con alguno de estos valores. **OBJETIVO:** Imputar los datos faltantes de una matriz con estructura de datos panel. **DATOS:** Los datos utilizados corresponden a 33 países, con el coeficiente de Gini registrado desde el año 2000 hasta el 2016. Es decir, 33 individuos y 17 variables. En total son 561 casos, de los cuales 135 (24%) son datos faltantes. **METODOLOGÍA:** Se opta por un modelo de estimación de datos faltantes que está pensado específicamente para estructuras de series de tiempo en tablas cruzadas. Esta propuesta incluye un modelo de estimación que considera los cambios en los individuos y las tendencias a lo largo del tiempo simultáneamente. En este trabajo, se asume que la probabilidad de datos faltantes en cualquier país es la misma, o en todo caso, que la probabilidad de datos faltantes de cualquier año es la misma. **RESULTADOS:** Se configuraron dos modelos de imputación considerando una sola variable (índice de Gini). El primer modelo fue sin restricciones, con 5 iteraciones. El segundo, un modelo de imputación flexible con el objetivo de mejorar los resultados. El primer modelo devuelve resultados que dejan en evidencia que las densidades de la imputación y de los datos reales no coinciden. En los datos observados se muestra la presencia de dos poblaciones, una que contempla a países con el índice de Gini entre los 20 y 40 puntos (en escala de 1 a 100) y otra población de países con el índice de Gini en escala de 40 a 70. En el caso de los datos imputados, se presenta también una estructura de dos poblaciones, sin embargo, se observa un claro desfase entre ambas densidades. El segundo modelo incorpora restricciones (información a priori). Con la introducción de la información a priori, el ajuste de la distribución del modelo de estimación es más adecuado. Además, se tienen intervalos de confianza, para las estimaciones de los datos faltantes, con menor amplitud que en el modelo sin restricciones. **CONCLUSIÓN:** El modelo con restricciones funciona mejor para la imputación de datos en el caso en estudio, y por lo tanto, las estimaciones son más confiables que con el modelo sin restricciones.

Palabras Clave: datos faltantes, datos panel, imputación múltiple

¹e-mail: ddelcallejo@uv.mx, mcanal@uv.mx, evernazza@iesta.edu.uy, alar.urruticoechea@ucu.edu.uy, ramon@iesta.edu.uy

Inclusión financiera en países de América del Sur con modelos de ecuaciones estructurales generalizadas

Hernández, María Paz y Caro, Norma Patricia¹

Facultad de Ciencias Económicas - Universidad Nacional de Córdoba

La inclusión financiera, hace referencia a que tanto personas físicas como jurídicas, tengan acceso a productos financieros útiles y asequibles que satisfagan sus necesidades en términos de transacciones, pagos, ahorros, créditos y seguros, y que a su vez sean prestados de manera responsable y sostenible. El estudio de la inclusión financiera y sus determinantes es un tema de suma importancia a la hora de pensar estrategias para mejorar el nivel de vida de un país. La inclusión financiera vía el desarrollo económico asociado, llevan a una reducción en los niveles de pobreza, logrando crecimiento en los sectores productivos. Es necesario llevar a cabo proyectos y/o políticas públicas focalizadas en estos aspectos. Esto motivó el estudio de los factores determinantes de la inclusión financiera en los países de ingresos medios – altos de América del Sur (Argentina, Brasil, Colombia, Ecuador y Perú, Chile y Uruguay). Para ello se utilizó la base de datos “Global Findex”, encuesta efectuada por el Banco Mundial de forma trienal en más de 140 países.

Se aplicó un modelo de ecuaciones estructurales generalizadas en el que se definen tres constructos (variables latentes) con sus indicadores: Acceso, Uso Digital y Educación Financiera y se establecen relaciones entre ellos, dado el marco teórico subyacente. Los resultados obtenidos indican que para el constructo Acceso (uso tradicional), los indicadores (tener cuenta financiera, crédito financiero y ahorro financiero) resultaron significativos en sentido positivo. Para el constructo Uso Digital, los indicadores e-commerce, tener cuenta móvil y el pago de servicios públicos mediante canales formales (distintos al efectivo) resultaron significativas, en sentido positivo. Para el constructo Educación Financiera, los indicadores de nivel educativo y nivel de ingresos, clasificados en altos y bajos, aportan positivamente al constructo. Se concluye que el acceso a los servicios financieros y la educación financiera afectan al uso digital.

Se pueden llevar a cabo iniciativas que promuevan la digitalización en el acceso y uso de servicios financieros, mejorando así la eficiencia del sistema financiero además de que se dejarían de gastar recursos en el abastecimiento de efectivo y en costos transaccionales, entre otros. Esto repercutirá en mayores canales de acceso, además de un abanico más amplio de servicios y productos financieros para la población, generando así mayor inclusión financiera. También es importante hacer foco en la educación financiera ya que es clave para lograr la sinergia con el acceso y uso de los servicios financieros.

Palabras Clave: inclusión financiera, modelos de ecuaciones estructurales generalizadas, países de América del Sur

¹e-mail: pacaro@unc.edu.ar

Mejores índices de similitud para datos binarios: evaluación sobre datos de comunidades vegetales

Patricia Susana Torres¹, Marta Beatríz Quaglino², Sergio Camiz^{3,1}

¹*Cátedra de Ecología – Facultad de Ciencias Agrarias – Universidad Nacional de Rosario (Argentina),*

²*Escuela de Estadística – Facultad de Ciencias Económicas y Estadística – Universidad Nacional de Rosario (Argentina),*

³*Istituto di Scienze del Patrimonio Culturale – Consiglio Nazionale delle Ricerche – Roma (Italia)*

El estudio se enfoca en la comparación de índices de similitud para datos de presencia/ausencia, un tema de particular interés en el área de Ecología de comunidades, donde la presencia de unidades taxonómicas en las unidades muestrales se considera más confiable que su frecuencia, sin pérdida significativa de información (Camiz, Torres, Pillar, 2017). El objetivo de este trabajo es examinar el comportamiento de los mejores índices de similitud, seleccionados por Torres (2019) sobre tres tablas de datos. Los índices seleccionados son doce y fueron elegidos de un total de 76, los cuales fueron examinados en el cumplimiento de 15 propiedades (11 de ellas ya siendo propuestas por Hubálek, 1982) que consideramos necesario que un índice debería cumplir para ser utilizado valiosamente.

Para esto, se consideran tres tablas de datos, representando una coenoclina y un coenoplano simulados, que responden a uno y dos gradientes ambientales hipotéticos respectivamente, y una transecta de datos de vegetación real. A las matrices de similitud resultado de la aplicación de los índices a las tres tablas, se les aplicó la descomposición propuesta por Gower (1966), análoga al Análisis de Componentes Principales. La comparación entre resultados se hizo para cada tabla a través de Análisis Procrusteano Generalizado (Gower, 1975), y en el caso de datos simulados, se compararon también las coordenadas resultantes con las coordenadas originales tal como fueron simulados los datos.

No sorprende encontrar que los índices de Jaccard y Sorensen son adecuados para analizar datos de comunidades vegetales, sin embargo se descubren otros índices que también dan resultados satisfactorios y no son tan conocidos en el área: Van der Maarel, Braun-Blanquet y Lamont & Grant. Todos estos índices no consideran en su definición las ausencias conjuntas, lo cual corrobora las declaraciones de algunos autores que destacan el mejor desempeño de estos índices en los estudios de Ecología de comunidades.

Palabras Clave: índices de similitud; ecología de comunidades; análisis de componentes principales, análisis procrusteano generalizado

¹e-mail: patrizia662@gmail.com, mquaglino@fcecon.unr.edu.ar, sergio@camiz.net

Análisis Factorial Múltiple Dual Mixto: propuesta metodológica y validación mediante una aplicación agronómica

Del Medico Ana P.¹, Vitelleschi María S., Pratta Guillermo M., Lavalle Andrea L.

Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística, Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario

*Cátedra de Genética, Facultad de Cs. Agrarias, Universidad Nacional de Rosario
IICAR, CONICET*

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

En este trabajo se presenta una propuesta metodológica, denominada Análisis Factorial Múltiple Dual Mixto (AFMDMix), para el análisis exploratorio de diferentes conjuntos o grupos de individuos que han sido descriptos por diferentes tipos de variables, cuantitativas y cualitativas. La técnica propuesta toma elementos del Análisis Factorial Múltiple Mixto (AFMMix) al tener en cuenta variables de distinta naturaleza y del Análisis Factorial Múltiple Dual (AFMD), al considerar una partición del conjunto de individuos. El AFMDMix posibilita estudiar comparativamente y en forma simultánea, las estructuras de asociación entre variables categóricas y las de correlación entre variables cuantitativas, inducidas por los grupos de individuos, así como las relaciones entre ambos grupos de variables.

La técnica de AFMDMix consiste en construir una matriz Z que resulta del preprocesamiento de varios grupos de individuos descriptos por un mismo conjunto de variables cuantitativas y cualitativas. A esa matriz Z se le realiza un ACP no normado que permite obtener la configuración de individuos y las variables y modalidades consenso. Luego, mediante proyecciones adecuadas, se obtienen las variables y modalidades parciales. De esta manera es posible analizar si la estructura de correlación o asociación entre variables, y entre tipos de variables, cambia de un grupo de individuos a otro.

Se aplicó el AFMDMix sobre conjuntos de poblaciones de germoplasma nativo de quínoa (*Chenopodium quinoa* Willd.) del Noroeste Argentino. Se tomaron dos muestras de germoplasma con diferentes orígenes -uno proveniente de Valles Secos (siete accesiones) y el otro de Valles de Altura (cinco accesiones)- que habían sido evaluados en base a variables morfológicas y agronómicas, 10 de ellas cuantitativas y cuatro cualitativas.

Las dos primeras dimensiones del análisis efectuado explicaron el 66,76 % de la variabilidad total de los datos. De la representación de las variables cuantitativas se observó que las estructuras de correlaciones presentaron las mayores diferencias entre las poblaciones provenientes de distintos ambientes para las variables días desde siembra a botón floral, longitud y ancho de hoja y longitud del glomérulo. Esto se corresponde con los coeficientes de correlación de Pearson calculados entre los pares de variables cuantitativas para cada grupo de individuos. Por otro lado, de la representación de las categorías de las variables categóricas se observó que las estructuras de asociación difirieron entre los dos grupos estudiados para las variables color de las estrías y color de la panoja, mientras que para las variables presencia de axilas y forma de la panoja fueron similares. Esto se corresponde con los coeficientes de asociación V de Cramer calculados entre cada par de variables categóricas para grupo de individuos. Por otro lado, se evidencia que en los Valles de Altura, valores altos de las variables días desde siembra a botón floral y días desde siembra a floración se asocian con panojas amarantiformes en tanto que valores bajos se asocian con panojas glomeruladas. Mientras que en los Valles Secos, las panojas glomeruladas se asocian con valores altos de diámetro de tallo y altura de planta, y las panojas amarantiformes se asocian con valores bajos de dichas variables.

En este trabajo se logró, a través de una propuesta original, estudiar comparativamente y en forma simultánea, las estructuras de asociación entre las variables categóricas y las de correlación entre las variables cuantitativas, inducidas por los grupos de accesiones de quínoa nativa del Noroeste Argentino, así como las relaciones entre ambos grupos de variables.

Palabras Clave: análisis factorial múltiple, grupos de individuos, variables cuantitativas y cualitativas, quínoa nativa

¹e-mail: adelmedico@fcecon.unr.edu.ar

Invarianza multirregional de la pobreza multidimensional en Argentina

Denise A. Gutiérrez Montecino, Adrián M. Moneta Pizarro¹

*Departamento de Estadística y Matemática, Facultad de Ciencias Económicas,
Universidad Nacional de Córdoba*

El fenómeno de la pobreza multidimensional y su medición en Argentina son temas que despiertan cada vez mayor interés. Los antecedentes de investigación disponibles se dividen en aquellos que adoptan el enfoque axiomático basado en la metodología de Alkire-Foster, que suponen predefinidas a las dimensiones, y los que se preocupan más por la identificación de las dimensiones mediante técnicas estadísticas. Dentro de este segundo grupo, algunos trabajos recientes muestran avances en la aplicación de métodos de análisis factorial robusto para lidiar con las características binarias de los datos disponibles, pero sin indagar aún sobre la existencia de diferencias regionales.

En este trabajo se analiza la invarianza factorial transversal de la pobreza multidimensional entre las regiones geográficas de Argentina. Para esto, primero se aplican técnicas robustas de análisis factorial exploratorio (AFE) y confirmatorio (AFC) sobre los microdatos de la Encuesta Permanente de Hogares (EPH) del cuarto trimestre del año 2019 a nivel global. Luego, el modelo identificado se somete a un análisis de invarianza factorial entre las cinco regiones geográficas del país.

Los resultados confirman la presencia de tres dimensiones que subyacerían a los indicadores de pobreza en Argentina y que son asociadas a factores de capacidad económica, infraestructura de la vivienda y entorno sociodemográfico. El nivel máximo de invarianza multirregional alcanzado fue el de invarianza débil, nivel que indica que la estructura dimensional y las cargas factoriales son estables entre regiones.

Estos resultados implican que no habría diferencias regionales significativas en cuanto a cuáles serían las dimensiones de la pobreza en Argentina; tampoco en los pesos que podrían utilizarse para ponderar los indicadores que se utilicen en la construcción de índices sintéticos. No obstante, son una advertencia respecto a posibles diferencias entre los interceptos, como así también entre las varianzas y covarianzas implicadas en las mediciones.

Palabras Clave: POBREZA MULTIDIMENSIONAL, ANÁLISIS FACTORIAL, INVARIANZA MULTIRREGIONAL, ARGENTINA.

¹e-mail: gutierrez.denise.5@gmail.com, amoneta@eco.uncor.edu

Estudio multidimensional de la pobreza mediante Análisis Factorial

Laura Nalbarte, Maximiliano Saldaña¹

Instituto de Estadística, Facultad de Ciencias Económicas y de Administración, Universidad de la República

Se aplica la metodología de Análisis Factorial para el estudio de los distintos factores que condicionan la pobreza y se analiza cómo se corresponde la pobreza subjetiva (autopercebida) con el status de pobreza, que surge de la metodología del cálculo de la línea de pobreza según método del ingreso (*Estimación de la Pobreza por el Método del Ingreso*, INE, 2018).

Los datos utilizados provienen de la *Encuesta Nacional de Gastos e Ingresos de los Hogares para 2016-2017*, efectuada en el Uruguay por el Instituto Nacional de Estadística. En dicha encuesta se relevó información sobre el consumo e ingreso de los hogares del país, junto a información referente a sus características socio-demográficas, materiales y de autopercepción de la pobreza. Se cuenta con datos muestrales que surgen un diseño muestral complejo correspondientes a 6889 hogares del territorio nacional.

Para tratar el objeto de estudio inicialmente se hace uso de técnicas de Análisis Factorial, en particular el Análisis de Correspondencias Múltiples (ACM), dado el carácter cualitativo de la mayoría de las variables empleadas, buscando así estudiar las relaciones entre las variables. Debido al considerable número de variables analizadas se efectúan varios ACM, cada uno considerando un bloque de variables que refieren distintas dimensiones, con la intención de no solo observar las relaciones dentro de dichos bloques sino también la que tienen con la pobreza subjetiva y con el status de pobreza del hogar. Las áreas que están representadas en bloque de variables son: elementos de confort del hogar, posesión y uso de tecnologías de la información, características de la vivienda, características socio-demográficas y pobreza subjetiva. Con el Análisis de Correspondencias Múltiples se tiene la intención de reducir el número de variables y resumir la información que estas presentan en índices para así considerar las distintas dimensiones de la pobreza en un Análisis Factorial Múltiple (AFM). Esta creación de un índice de pobreza multidimensional está inspirado en el desarrollado en *Pobreza Multidimensional Utilizando Análisis Factorial* (Altmark, Massa y Nalbarte, 2014). Para el análisis de los datos se empleó el software estadístico R (R Core Team, 2020) a través de la librería *FactomineR* (Husson et al, 2008).

Un mayor número de hogares se autorreportan como pobres que aquellos que están por debajo de la línea de pobreza. A pesar de esta brecha, tanto la pobreza subjetiva como el status de pobreza tienden a asociarse con el hecho de no disponer de ingresos suficientes para satisfacer necesidades básicas. A su vez estas tipologías de pobreza tienden a asociarse con la carencia de elementos de confort, de dispositivos como los celulares y con problemas en la materialidad de la vivienda. Los resultados preliminares muestran la existencia de un índice de pobreza multidimensional sintetizado en un solo puntaje factorial u score al aplicar el AFM sobre los primeros factores de cada ACM por bloques, el que se asocia con la pobreza tanto subjetiva como la definida por el método del ingreso. El equipo de investigación se propone corroborar los resultados analizando los datos más en profundidad y complementándolos con diferentes métodos de clustering.

Palabras Clave: Análisis Multivariado, Análisis Factorial de Correspondencias Múltiples, Análisis de Clustering, Pobreza, Pobreza Subjetiva.

¹e-mail: maximiliano.saldana@fcea.edu.uy

Determinantes de un historial completo de aportes al sistema previsional argentino

Milva Geri, Fernanda Villarreal ¹

Instituto de Matemática (INMABB), Departamento de Matemática, Universidad Nacional del Sur (UNS)-CONICET, Bahía Blanca, Argentina

Instituto de Investigaciones Económicas y Sociales del Sur (UNS-CONICET); Departamento de Economía (Universidad Nacional del Sur), Departamento de Matemática (Universidad Nacional del Sur), Bahía Blanca, Argentina.

Uno de los factores que pueden incidir sobre el desempeño de un sistema previsional es la capacidad del mercado laboral para generar puestos de trabajo formales en forma sostenida. A lo largo de la historia del sistema previsional argentino se ha recurrido repetidas veces a la flexibilización de las llamadas condiciones de elegibilidad que consisten en alcanzar una edad mínima y un mínimo de años de aporte. El objetivo de este trabajo es identificar los determinantes de la probabilidad de que un individuo que aportó al menos una vez al sistema previsional argentino cuente con un historial completo de aportes al alcanzar la edad mínima jubilatoria. La muestra empleada proviene de declaraciones juradas de empleadores, del padrón de empresas de la Administración Federal de Ingresos Públicos (AFIP) y de la Administración Nacional de la Seguridad Social (ANSES) para el periodo enero 1996 -diciembre de 2015 publicada por el Ministerio de Trabajo, Empleo y Seguridad Social en el año 2018. Las variables demográficas con las que cuenta la base son el sexo y la fecha de nacimiento. Las variables laborales se asocian al puesto de trabajo y son la provincia y rama de actividad donde la relación laboral tiene lugar, así como el tamaño y la antigüedad de la empresa y la remuneración que percibe el individuo. Dado que la unidad observacional es la relación laboral, la matriz de datos original tiene cerca de 1,4 millones de filas (una para cada relación laboral registrada). A partir de un algoritmo en R se obtuvo una nueva matriz de datos cuya unidad observacional es el individuo que cumple con la edad mínima jubilatoria en 2015: varones nacidos en 1950 o antes y mujeres nacidas en 1955 o antes. Luego de eliminar observaciones que presentaban valores perdidos, se obtuvo información completa acerca de 31.488 individuos. Se construyó la variable dicotómica “Completo Aportes” o “No Completo Aportes” al momento de jubilarse y se propuso un modelo de Regresión Logística Binaria con el objetivo de obtener una estimación ajustada de la probabilidad de que un individuo cuente con un historial completo de aportes al momento de jubilarse en función de las variables sexo, edad del individuo, total de relaciones laborales protagonizadas, remuneración máxima percibida en el período y de variables asociadas a la relación laboral de mayor ingreso: la provincia y la rama de actividad donde dicha relación tenía lugar y el tamaño y antigüedad de la empresa contratante. Se utilizó el software SPSS versión 25 para obtener los estimadores a partir de una regresión logística. Entre las conclusiones se destaca que una persona de sexo masculino tiene casi dos veces más probabilidad de completar sus aportes al momento de jubilarse que una persona de sexo femenino, a pesar de controlar por otros factores también relevantes como la provincia donde tiene lugar la relación laboral, la rama de actividad, el tamaño y la antigüedad de la empresa y el número de relaciones laborales protagonizadas. Se espera contribuir de esta manera al debate acerca de los determinantes del cumplimiento de una de las condiciones de elegibilidad para acceder a los beneficios previsionales que actualmente consiste en reunir 30 años de aportes. Los resultados del trabajo podrían ser de interés para el diseño de políticas enfocadas a los individuos que presentan menores probabilidades de cumplir con este requisito.

Palabras Clave: regresión logística - sistema previsional -aportes

¹e-mail: milvageri@gmail.com, fvillarreal@uns.edu.ar

ANÁLISIS DE UN PANEL DE CONSUMIDORES DE YERBA MATE UTILIZANDO METODOLOGÍA MULTIVARIADA

FERNÍCOLA M.^a , DRUNDAY F.^b , CALVIÑO A.^c , y NUÑEZ M.^{a1}

^a *Cátedra de Matemática*, ^b *Cátedra de Fisiología*, ^c *IQUIMEFA, UBA-CONICET*

Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires, Argentina

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 a *Ilex paraguariensis* dando lugar a la categoría de Yerba mate compuesta, y esto produce un sabor diferente al mate tradicional, muy aceptado por los consumidores. La yerba mate presenta un elevado índice de consumo doméstico y también se diversifica su empleo en numerosos mercados a nivel mundial (Bracesco y col., 2010, Filip, 2011). Las características del olor y la apariencia de la yerba mate son atributos claves en la aceptación del consumidor. El objetivo de este trabajo es evaluar la reproducibilidad y concordancia de un panel de consumidores no entrenado, que realizó, mediante perfiles libres, la descripción de las similitudes y diferencias sensoriales de un conjunto de yerbas comerciales, elaboradas y compuestas. En el caso del Perfil de Libre Elección (PLE), hay total libertad para la elección del criterio y la selección de los descriptores y no requiere una puesta en común del mismo, a diferencia del abordaje tradicional del Análisis Sensorial (disciplina científica que se ocupa de la caracterización y medición de los atributos de un producto alimenticio que son percibidos por los sentidos), donde se obtiene esta caracterización a través de un panel de expertos. Esta metodología requiere de una larga fase de formación, entrenamiento y puesta en común del vocabulario del panel profesional. Par ello se consideraron 18 marcas comerciales de yerba mate y yerba mate compuesta a ser evaluadas por un grupo de 20 consumidores no entrenados, mediante el esquema de PLE en tres sesiones experimentales. Cada panelista elabora una lista de descriptores que le permite caracterizar el producto. Debido a la variedad de los términos seleccionados, se estudian los vocablos con los que la mayoría de los consumidores coinciden, dos de apariencia, y uno de aroma. Se presentan los resultados de un atributo de apariencia y otro de olor de las yerbas evaluadas. Se evaluaron las características de apariencia y olor que incluyeron yerbas despaladas, con palo y compuestas con agregado de hierbas como peperina o menta. Los datos de perfil libre fueron analizados mediante Análisis de Procrustes Generalizado (Dijksterhuis y Punter, 1990). El análisis estadístico abarcó la aplicación de análisis de varianza de dos factores (yerba y sesión), análisis por componentes principales (ACP) y Análisis de Clasificación Jerárquica. La clasificación jerárquica nos permitió detectar la reproducibilidad o capacidad discriminatoria de los jueces. La localización de los jueces en el plano permite encontrar grupos afines, de modo que, jueces cercanos perciben el atributo en forma similar y jueces alejados entre sí evalúan el atributo en forma diferente. Se comprobó la existencia de diferencias significativas entre yerbas, para cantidad de palo y para nota menta-mentol. No hubo diferencias entre las sesiones para ambos atributos, lo cual muestra la consistencia de los jueces y también se observa que la interacción yerba x sesión no resultó significativa. El ACP permitió la localización de grupos afines de jueces, de modo que, jueces cercanos se perciben como similares y jueces alejados entre sí como diferentes. En líneas generales se concluye que el panel de consumidores habituales de mate resultó consistente en sus evaluaciones de atributos de apariencia y olor de un conjunto de yerbas comerciales. La evaluación de olor, sin embargo, resultó una tarea más compleja para algunos evaluadores.

Palabras Clave: Perfil de Libre Elección, Análisis Multivariado, Análisis Sensorial, Yerba Mate Compuesta

¹e-mail: mfernocola21@gmail.com

Tema 17: Salud

Un estudio de la relación entre nuevos casos de COVID-19 y cantidad acumulada de vacunados en Argentina

Lila Ricci¹, Diana Kelmansky² ¹

¹Centro Marplatense de Investigaciones Matemáticas, Facultad de Ciencias Exactas y Naturales, Universidad Nacional de Mar del Plata; ²Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires

En el marco de la pandemia de COVID-19 en Argentina, este trabajo analiza en forma retrospectiva el efecto de la vacunación sobre los siguientes eventos: contagios, internaciones en terapia intensiva y fallecimientos. El estudio se basa en datos públicos provistos por el Ministerio de Salud y abarca desde el comienzo de la vacunación en diciembre de 2020 hasta mediados de 2021. Dado que la campaña de vacunación se llevó a cabo primero en pacientes de más de 80 años, incorporándose paulatinamente a los otros grupos etarios, el análisis se hace en forma estratificada dentro de cada grupo. Se describe la evolución de la cantidad de eventos semanales y de la cantidad acumulada de primeras dosis y de segundas dosis aplicadas; los resultados se expresan como porcentajes del correspondiente valor máximo, para cada evento y para cada grupo etario. Se compara la cantidad de eventos al inicio de la campaña de vacunación con dichas cantidades al finalizar el estudio. Se estudia también la asociación entre la cantidad de eventos semanales y la cantidad acumulada de vacunas aplicadas hasta esa semana.

Palabras Claves: Casos de COVID-19, Vacunación, Dinámica epidemiológica, Argentina

¹e-mail: lricci@mdp.edu.ar, dkelman@ic.fcen.uba.ar

Psoriasis y su impacto metabólico

María de Luján Calcagno¹ Silvina Del Duca² Jimena Soutelo³

Facultad de Farmacia y Bioquímica, Departamento de Fisicomatemática, Universidad de Buenos Aires

Servicio de Endocrinología, Hospital Churrucá-Visca

La psoriasis es una enfermedad crónica inflamatoria sistémica, asociada a múltiples comorbilidades, en especial en su forma severa. El estado proinflamatorio es el nexo con las enfermedades concomitantes, como resistencia a la insulina, diabetes, dislipemia, obesidad, hipertensión, enfermedad cardiovascular (ECV) y alteraciones del metabolismo mineral óseo (MMO).

El objetivo del presente trabajo es evaluar si los pacientes con psoriasis presentan factores de riesgo de ECV y deterioro del MMO. Es un trabajo prospectivo, observacional, transversal. Se evaluaron pacientes con psoriasis (P) y controles sanos (C), atendidos en el Servicio de Endocrinología del Hospital Churrucá-Visca. Se excluyeron pacientes con antecedentes de ECV, diabetes, enfermedad renal y osteoporosis, entre otras. Se recabaron antecedentes personales, hábitos y medicación habitual. Se realizaron medidas antropométricas, con cálculo del índice de masa corporal (IMC) y toma de presión arterial. Se evaluó la psoriasis de acuerdo al Índice de Severidad del Área de Psoriasis (PASI). Se solicitó hemograma, Velocidad de Sedimentación Globular (VSG), hepatograma, glucemia, creatinina, insulina, perfil lipídico y metabolismo fosfocálcico. Se calculó el filtrado glomerular (CKD-EPI), el índice HOMA y el índice TG/HDL. Se efectuó eco Doppler carotideo para evaluar grosor de íntima media y presencia de placas. Asimismo, se realizó densitometría ósea (DMO) de columna lumbar y cadera izquierda y se interrogó sobre fracturas por fragilidad.

Se realizó un análisis estadístico descriptivo detallado y se aplicaron, entre otros, el test t de Student, test de Mann-Whitney, ANOVA de un factor fijo, test a posteriori de Bonferroni y test de Kruskal-Wallis. Los resultados obtenidos muestran que no hubo diferencia significativa en la edad, ni en la distribución de género. Los pacientes con psoriasis presentaron mayor peso, IMC y VSG que los controles. Dentro del perfil lipídico, solo se halló diferencia significativa en HDL. También se halló diferencia en la calcemia. Posteriormente se dividió a los individuos analizados de acuerdo a la severidad de la enfermedad psoriásica: controles (C), psoriasis leve (PL) y psoriasis moderada-severa (PMS). Se halló diferencia significativa en la calcemia y en el índice TG/HDL, pero no así en el índice HOMA. Por otra parte, en los pacientes con psoriasis se evaluó la presencia de placa ateromatosa. Al comparar ambos grupos, aquellos con placa tuvieron mayor presión arterial sistólica ($p=0,009$), peor FG y mayor grosor de íntima media. No hubo diferencias significativas en DMO ni presencia de fracturas por fragilidad.

Las conclusiones a las que arribamos contemplan el hecho de que existe una base fisiopatológica que justifica una asociación entre ECV, osteoporosis y psoriasis. La producción excesiva de citoquinas proinflamatorias, sería el mecanismo común. Numerosas revisiones reportaron a la psoriasis como un factor de riesgo independiente para ECV. Los estudios sobre afectación de la DMO y aumento del riesgo de fractura, son limitados y contradictorios. En nuestra experiencia, hallamos que los pacientes con psoriasis presentaron mayor peso, IMC e índice TG/HDL como marcador de resistencia a la insulina. Mas aun, estos parámetros fueron peores en aquellos con placa carótida. Con respecto al MMO, solo hallamos una disminución de la calcemia, que correlaciona con la severidad de la psoriasis. Está justificado realizar otros estudios con un enfoque multidisciplinario, para definir potenciales comorbilidades.

Palabras Clave: Psoriasis, Enfermedad Cardiovascular, Metabolismo Mineral Óseo

¹e-mail: mariadelujancalcagno@gmail.com

²e-mail: silvinadelduca@gmail.com

³e-mail: mjimenasoutelo@gmail.com

Desigualdades en la realización de actividad física en adultos según gradiente socioeconómico en Argentina: un estudio a partir de la Encuesta Nacional de Factores de Riesgo 2018.

Rozenblit Valentín, Fernández María Soledad, Pérez Adriana¹

Grupo de Bioestadística Aplicada, Instituto de Ecología, Genética y Evolución de Buenos Aires; Instituto de Cálculo; FCEN, UBA-CONICET, Universidad de Buenos Aires

La realización de actividad física (AF) como hábito de vida es altamente beneficioso para la salud física y mental, siendo que el sedentarismo constituye un factor de riesgo para enfermedades no transmisibles. En Argentina, si bien la prevalencia en adultos de realización de AF media o alta (definida como al menos 3 días de actividad física intensa por 20 minutos cada semana) aumentó en los últimos años, pasando de 45.3 % en 2013 a 55.8 % en 2018, sigue siendo exigua comparada con las recomendaciones internacionales. Existe gran evidencia de que las diferencias en el nivel socioeconómico (NSE) de las personas se asocian con desigualdades en la capacidad de adoptar conductas saludables que, a su vez, contribuyen a las inequidades en salud. El objetivo de este trabajo es estudiar la relación entre la realización de AF y distintas dimensiones del NSE (educación, ingresos, bienes materiales) en adultos de Argentina. Se trabajó con la cuarta edición de la Encuesta Nacional de Factores de Riesgo 2018 (ENFR), conformada por una muestra de 29.224 individuos mayores de 18 años de 26 aglomerados urbanos de Argentina. Se implementaron regresiones logísticas para analizar la relación entre la realización de AF (media o alta) y NSE a escala individual y hogar: nivel educativo (hasta primario completo (P), secundario incompleto o completo (S), universitario incompleto o completo (U)), quintil de ingreso y presencia de carencias materiales y de vivienda (CMV). Esta última es una variable construida a partir del tipo de vivienda, el material del suelo y del cielorraso, tipo de baño, hacinamiento y tipo de suministro de agua. Se incluyeron como variables de control al género, el rango etario y la provincia. Se evaluaron posibles interacciones entre las distintas dimensiones del nivel socioeconómico y el género. La prevalencia de AF fue mayor en adultos varones (V: 56.1 %) que en mujeres (M: 51.3 %) ($p < 0.0001$). A partir de los odds ratio (OR) se vio que tanto el nivel educativo como el quintil de ingreso se asociaron en forma directa con la realización de AF (Quintil de ingreso: $OR_{Q1-Q5} = 0.720$, $IC_{95\%}: [0.643; 0.805]$, Nivel educativo: $OR_{P-U} = 0.605$, $IC_{95\%}: [0.558; 0.656]$). Se halló asociación inversa con la presencia de CMV ($OR_{Sin\ CMV-Con\ CMV} = 1.12$, $IC_{95\%}: [1.06; 1.18]$). La prevalencia de AF disminuyó con la edad y fue menor para las mujeres ($OR_{M-V} = 0.822$, $IC_{95\%}: [0.783; 0.862]$). No se encontró interacción entre género y ninguna de las tres dimensiones de NSE ($p > 0.05$ en todos los casos). Utilizando una muestra representativa de la población adulta de Argentina, se halló relación entre la prevalencia de realización de AF y todas las dimensiones del NSE a escala individual y hogar analizadas. Las personas menos aventajadas socioeconómicamente fueron las que presentaron menor prevalencia de AF, independientemente del género y la edad. La ausencia de interacción entre género y NSE sugiere que las brechas entre varones y mujeres atraviesan todo el gradiente socioeconómico. Estos resultados aportan evidencia sobre las desigualdades socioeconómicas en la realización de AF en adultos de Argentina, que podrían explicar las desigualdades socioeconómicas en mortalidad y otros problemas de salud. Los grupos con menor NSE y las mujeres deberían ser prioritarios al momento de enfocar políticas de concientización y promoción de la AF de manera de reducir las desigualdades existentes.

Palabras Clave: Hábitos saludables, Modelos Lineales Generalizados, Grandes Bases de Datos, Datos Públicos, Regresión Logística

¹e-mail: valentin.rozenblit@gmail.com, sfernandez79@gmail.com, adrianaperez000@gmail.com

ANÁLISIS DEL TIEMPO DE ESTANCIA HOSPITALARIA EN PACIENTES CON COVID-19 COMO HERRAMIENTA PARA LA TOMA DE DECISIONES EN EL SECTOR SALUD

MARIA VIRGINIA PISANI, ALICIA QUINTANA, FERNANDA VILLARREAL,
MARÍA EUGENIA ELORZA¹

Dto. de Matemática, Universidad Nacional del Sur (UNS), Bahía Blanca, Argentina
Instituto de Matemática (INMABB), Dto. de Matemática, Universidad Nacional del Sur
(UNS)-CONICET, Bahía Blanca, Argentina
IIESS (UNS-CONICET), Dto. de Economía, Universidad Nacional del Sur (UNS), Bahía Blanca,
Argentina

La variable duración de la estancia hospitalaria (length of stay (LOS)) representa el tiempo que un paciente permanece internado en una institución hospitalaria y constituye un importante indicador para la gestión de recursos hospitalarios y la planificación de los servicios de salud. En las internaciones asociadas a los procesos de atención de pacientes con COVID, el comportamiento de la variable tiempo de estancia, brinda información de utilidad para la toma de decisiones que mejoren la gestión de los recursos hospitalarios afectados en estos procesos de atención, entre ellos el recurso cama. Por este motivo, es sumamente relevante profundizar en el análisis estadístico de la variable mencionada.

El objetivo de este trabajo es estudiar la distribución del tiempo de estancia de pacientes internados por COVID-19 positivo en una institución hospitalaria pública y la búsqueda de factores que afectan a su distribución. Se analizó una muestra de 264 pacientes con COVID positivo que resultaron internados en un hospital público de una ciudad ubicada a 500 km de la ciudad autónoma de Buenos Aires perteneciente a la provincia de Buenos Aires entre agosto de 2020 y julio de 2021.

Se realizó en primer lugar, el diseño de la base de datos (construcción de variables y depuración de casos) para luego continuar con un análisis exploratorio-descriptivo y, posteriormente, un análisis inferencial utilizando, entre otros, 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 variables sociodemográficas, edad y género, y variables que aportan información clínica del paciente como antecedentes y motivos de la consulta, considerando que ambos tipos de variables pueden afectar el tiempo de permanencia en el hospital.

Entre los principales resultados hallados cabe mencionar: 1) la distribución del tiempo de internación presenta una asimetría positiva con una mediana de 5 días. Se encontraron: 2) diferencias significativas entre las distribuciones del tiempo de internación según la edad del paciente, observándose que la permanencia en el hospital fue mayor para los casos de pacientes con al menos 60 años 3) diferencias significativas para la variable en estudio de acuerdo a si la internación se produjo en Terapia Intensiva o Clínica General-Medicina General 4) evidencias estadísticas significativas de la dependencia entre las variables edad y antecedentes (presencia/ausencia). No se encontraron: 5) diferencias significativas en el tiempo de internación de acuerdo al sexo ni a la cantidad de síntomas presentes en el paciente al momento de concurrir al hospital

Los resultados obtenidos constituyen los de un estudio preliminar, para futuros análisis que contarán con información de más casos y variables. Sus conclusiones, en conjunto, se presentan como potenciales para la gestión hospitalaria y el diseño de políticas, dado que el mejoramiento de la calidad de la información permite un proceso de toma de decisiones más eficiente, al momento de asignar recursos y estimar los costos asociados a este problema de salud pública.

Palabras Clave: Estadística Inferencial- Tiempo de estancia- COVID- 19- Gestión hospitalaria

¹e-mail: mvpisani@uns.edu.ar, quintana@uns.edu.ar, fvillarreal@uns.edu.ar, meugilorz@hotmail.com

Construcción de un test diagnóstico sobre la educación de la sociedad argentina frente a la enfermedad COVID-19.

Janina Micaela Roldan, María Cristina Martín¹

Facultad de Ciencias Exactas y Naturales, Universidad Nacional de La Pampa

Departamento de Matemática, Universidad Nacional del Sur

El contexto de pandemia generado a partir de la enfermedad COVID-19 ha obligado a los países a tomar medidas que contribuyan a mermar la propagación de la enfermedad. En este sentido, el aislamiento social, preventivo y obligatorio decretado en la Argentina, como en varios países, ha sido una de las medidas efectivas para disminuir la cantidad de casos positivos de COVID-19 diarios. No obstante, es claro que el confinamiento extremo no puede durar demasiado tiempo ya que afecta a las personas, tanto individual como colectivamente, produciendo desde afecciones económicas hasta psicológicas. Es importante, por lo tanto, suprimir restricciones a la sociedad para que en el menor tiempo posible se vuelva a un estado, al menos parecido, a cómo se vivía previo a la pandemia mundial, pero con ciertos cuidados debido a que la enfermedad COVID-19 aún no se ha erradicado, y seguramente, demorará algunos años en hacerlo.

Por lo expuesto, es esencial evaluar el comportamiento de los ciudadanos argentinos en relación a buenas prácticas que contribuyan a evitar contraer la enfermedad en cuestión. Sabiendo que los cuestionarios (test) son los instrumentos de medida más utilizados para la obtención de datos a fin de realizar un diagnóstico sobre el comportamiento de las personas, resulta fundamental valerse de un test para cuantificar, en alguna escala de medida, la educación de los argentinos frente al COVID-19. Sin embargo, al día de hoy no se cuenta con tal instrumento de medida y este es un aspecto que no ha sido considerado en las líneas de investigación relacionadas al COVID-19, principalmente, porque la construcción de un instrumento de medida es un proceso complejo que está conformado por varias etapas que se suceden en el tiempo, y siguen estándares de calidad con el fin de validar las inferencias que se realizarán a posteriori a partir de las puntuaciones obtenidas.

Este trabajo tiene como objetivo presentar los primeros aspectos que conforman la construcción de un test útil para evaluar la educación de los argentinos frente a la enfermedad COVID-19. Estas fases consisten en la delimitación del marco general, en la definición de la variable a medir y otras variables clasificatorias, en especificaciones relacionadas con el instrumento de medida y, fundamentalmente, en la construcción de los ítems.

Palabras Clave: COVID-19 – Educación– Sociedad Argentina - Test – Ítems

¹e-mail: roldan.janina@exactas.unlpam.edu.ar, maritamartin11@gmail.com

Hábitos alimentarios e inequidades socioeconómicas: Un análisis del consumo de frutas y verduras en adultos a partir de datos públicos de Argentina

Safarian María Jazmín^a, Bruveris Ana Paula^b, Pérez Adriana^a, Fernández María Soledad^{a1}

^aGrupo de Bioestadística Aplicada, Instituto de Ecología, Genética y Evolución de Buenos Aires; Instituto de Cálculo; FCEN, UBA-CONICET; ^bFacultad de Ciencia y Tecnología de los Alimentos, Universidad Nacional del Comahue

Una alimentación inadecuada constituye un factor de riesgo para enfermedades no transmisibles, las cuales constituyen la principal causa de morbimortalidad a nivel nacional (MSal, 2016). La Organización Mundial de la Salud (OMS) recomienda consumir cinco porciones diarias de frutas y verduras. Sin embargo, en Argentina se ha observado una reducción constante en el consumo de estos alimentos en los últimos 20 años, en particular en adultos evidenciado en las sucesivas ediciones de la Encuesta Nacional de Factores de Riesgo (ENFR). A su vez, existe fuerte evidencia de que la mayoría de los factores de riesgo de enfermedades no transmisibles están estructurados a lo largo del gradiente socioeconómico. Este gradiente se define en el contexto de las inequidades en salud, donde un deterioro en las condiciones de salud de las personas está asociado a su nivel socioeconómico (NSE). El objetivo de este trabajo es analizar las inequidades sociales en el consumo de frutas y verduras a partir de la última edición de la ENFR (2018). La muestra estuvo conformada por 29.224 individuos mayores de 18 años de 26 aglomerados urbanos de Argentina. Se consideró como consumo de frutas y verduras adecuado (CFV) a la ingesta de tres o más porciones promedio diarias según las Guías Alimentarias para la población Argentina (equivalentes a los 400 gramos recomendados por la OMS). Para modelar el CFV se aplicaron modelos lineales generalizados con distribución Bernoulli. Las variables predictoras de NSE consideradas a escala individual fueron el nivel educativo alcanzado (hasta primario completo (P), secundario incompleto o completo (S) y estudios superiores incompletos o completos (U)) y el quintil de ingresos. A escala hogar se incluyó la presencia de carencias materiales y de vivienda (CMV). Para esta última variable se consideraron características estructurales de la vivienda, como el material del suelo y del cielorraso, el tipo de baño, suministro de agua y nivel de hacinamiento. El modelo fue ajustado por género, rango etario (18-24, 25-34, 35-49, 50-64 y 65 años o más) y NBI provincial. Se evaluaron interacciones del género y del rango etario con las variables predictoras de NSE. Todas las variables de NSE se asociaron significativamente al CFV en la dirección esperada. La prevalencia de CFV (P [IC95%]) aumentó con el nivel educativo ($P_P = 0,163 [0,154; 0,172]$, $P_S = 0,184 [0,177; 0,192]$, $P_U = 0,258 [0,248; 0,269]$) y disminuyó en hogares con carencias materiales ($P_{Con\ CMV} = 0,191 [0,182; 0,200]$, $P_{Sin\ CMV} = 0,207 [0,201; 0,214]$). Se halló interacción significativa entre el género y el quintil de ingreso. La prevalencia de CFV aumentó según el quintil de ingresos pero en mayor medida en mujeres que en hombres ($P_{Q5-mujer} = 0,294 [0,277; 0,312]$, $P_{Q5-varón} = 0,208 [0,193; 0,224]$ vs $P_{Q1-mujer} = 0,165 [0,152; 0,179]$, $P_{Q1-varón} = 0,158 [0,142; 0,175]$). En relación a las variables de control, el CFV aumentó con la edad ($p < 0,001$) y disminuyó con el tercil de NBI provincial ($p < 0,001$). Los resultados de este trabajo aportan evidencia acerca de la asociación entre el gradiente socioeconómico, medido en múltiples dimensiones (ingresos, educación y bienes materiales) y un hábito saludable como es el CFV. Los resultados encontrados en relación a un mayor CFV para mujeres mostraron que la brecha con hombres aumenta con el NSE. En este sentido se destaca la necesidad de orientar políticas públicas para fomentar el CFV en poblaciones vulnerables pero considerando las diferencias de género existentes según el NSE.

Palabras Clave: Hábitos saludables, modelos lineales generalizados, grandes bases de datos

¹e-mail: safarianjazmin@gmail.com, anapaulabruveris@gmail.com, adrianaperez000@gmail.com, sfernandez79@gmail.com

INVESTIGACIÓN DEL HÁBITO TABÁQUICO EN PACIENTES ONCOLÓGICOS Y ALLEGADOS

DRA. NUÑEZ M.^a, FAR. BONADÍES E.^a, FAR. MONTENEGRO BRUSOTTI J. ^a,
DRA. VERZURA M.A.^b, DRA. MAUR PEROTTI J.A.^b, DR. PEREZ MESA P.^b,
DR. SPECTERMAN S.^b, DR. LASTIRI J.M.^{b1}

^a *Cátedra de Matemática, Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires*

^b *Hospital Italiano de Buenos Aires, Argentina*

Las enfermedades no transmisibles (ENT) son actualmente la principal causa de mortalidad mundial. Un alto porcentaje de ENT son prevenibles y comparten los mismos factores de riesgo. Los cinco principales riesgos para la salud están relacionados con el comportamiento y la alimentación: índice de masa corporal alto, bajo consumo de frutas y hortalizas, inactividad física, tabaquismo e ingesta excesiva de alcohol. Estos factores causan el 30 % de las muertes por cáncer. Las variables asociadas con el desarrollo y perpetuación del tabaquismo incluyen el nivel educacional, la precocidad en el inicio del tabaquismo, el consumo en el hogar o en grupo familiar, entre otras. Los objetivos de este trabajo son: a) conocer la prevalencia y características del tabaquismo, y la conciencia de daño en pacientes con cáncer (GP) y sus allegados (GF), a través de una encuesta realizada en el período 2016-2017. Comparar las características del hábito en pacientes con neoplasias asociadas y no asociadas. Comparar estos resultados con la población estudiada con la misma encuesta realizada en el período 1997 - 1998, y b) Conocer la repercusión que han tenido las nuevas leyes de protección contra el tabaco en la conciencia y prevalencia del hábito. Ambas encuestas se realizaron en el Hospital Italiano de Buenos Aires (HIBA) en los períodos 2016 - 2017, y 1997 - 1998. Se obtuvieron muestras de 364 pacientes y 366 familiares, y 350 y 390, respectivamente. Se registraron los siguientes datos: Sexo, Edad, Factores de riesgo: hábito de fumar, frecuencia de cigarrillos, tiempo que lleva fumando, Características del hábito, Conocimiento acerca de la enfermedad, Deseos de cesación, Conciencia de daño. La aplicación de metodología estadística multivariada como, Análisis Factorial de Correspondencias, Regresión Logística y Árboles de Clasificación, nos permitió tener una idea del comportamiento de todos los factores de riesgo en su conjunto. Algunas de las conclusiones fueron: Un tercio de la patología oncológica en nuestro centro se asocia al tabaquismo. En el caso del GF la frecuencia en el consumo es levemente inferior a la conocida en nuestro medio, y posiblemente relacionada con el mayor número de mujeres en este grupo. En nuestra serie los fumadores de más de 20 cigarrillos diarios fueron la población más asociada a neoplasias del fumador, y fue significativamente mayor en hombres, especialmente con neoplasias relacionadas. Los pacientes de la serie anterior no demostraron interés en la cesación del hábito, en cambio en la encuesta actual, hubo casi un 20 % menos de fumadores, así como alrededor de un 20 % más de ex fumadores, aunque no alcance una diferencia estadísticamente significativa. Tanto en el GP como en el GF el nivel de conocimiento acerca del daño que genera el tabaco ronda en el 90 %, demostrando una diferencia significativa con la población antes encuestada. Mediante el análisis factorial de correspondencias se observó que los factores predominantes corresponden al hábito de fumar y al conocimiento sobre la enfermedad y deben tenerse en cuenta a la hora de planificar medidas de prevención y/o detección temprana del cáncer. Con estos resultados, si bien se hallaron algunas diferencias que hablan del beneficio produjeron las acciones en la legislatura como en la difusión, consideramos que es fundamental seguir trabajando para que estas diferencias sean más notorias.

Palabras Clave: hábito tabáquico, noción de daño, métodos multivariados

¹e-mail: myriam@ffyba.uba.ar

Modelos lineales generalizados y de cuasiverosimilitud para la evaluación del funcionamiento renal

Pehuén Fernández*, Walter Douthat*, Carlos Chiurciu*, María Laura Nores^{†1}

**Servicio de Nefrología, Hospital Privado Universitario de Córdoba; Universidad Católica de Córdoba; Instituto Universitario de Ciencias Biomédicas de Córdoba.*

[†]*Facultad de Matemática, Astronomía, Física y Computación, Universidad Nacional de Córdoba.*

La enfermedad renal es un problema global de salud pública que afecta a más de 750 millones de personas alrededor del mundo. El conocimiento de la tasa de filtrado glomerular (TFG) juega un rol fundamental en la evaluación del funcionamiento renal y en la detección y seguimiento de la enfermedad. Muchas organizaciones recomiendan el uso de ecuaciones a partir de la concentración de creatinina sérica para estimar la TFG, de manera de facilitar la detección, evaluación y manejo de la enfermedad renal crónica (ERC). Diversas fórmulas se utilizan a diario en la práctica clínica para predecir la TFG, como CKD-EPI, MDRD-4, MDRD-6 y MCQ. El objetivo de este trabajo fue desarrollar una nueva ecuación basada en datos de Argentina y comparar su desempeño con las ecuaciones preexistentes.

Los datos utilizados corresponden a adultos residentes en Argentina a los cuales entre 2007 y 2017 se les midió la TFG con iotalamato (patrón de referencia) en el Hospital Privado Universitario de Córdoba, Córdoba, Argentina ($n = 583$).

Inicialmente se ajustaron modelos lineales transformando la respuesta (TFG) mediante la función logaritmo natural, como en las fórmulas de uso actual. Para preservar la escala original de la variable respuesta, se ajustaron modelos lineales generalizados con distribución Gamma y enlace logarítmico y modelos de cuasiverosimilitud, especificando una relación media-varianza lineal $V(\mu) = \mu$ y enlace logarítmico. Entre las variables predictoras, se seleccionaron modelos incluyendo concentración de creatinina sérica, edad, sexo y una variable que indica si el paciente es monorreño. También se consideraron fórmulas más complejas con albúmina sérica y urea.

Para estudiar la capacidad de predicción de las nuevas ecuaciones se utilizó validación cruzada de 10 iteraciones, separando al azar el total de observaciones en 10 grupos, dejando fuera uno por vez para ajustar el modelo y prediciendo a partir de ese ajuste los valores de TFG en el grupo excluido.

Las medidas utilizadas para evaluar el rendimiento de las nuevas ecuaciones y de las previas fueron $RMSE$, R^2 ajustado, sesgo (diferencia entre la TFG estimada y medida), P_{15} y P_{30} (porcentaje de observaciones cuya TFG estimada difiere de la medida en no más del 15 o 30% de la TFG medida, respectivamente). Por otra parte, la ERC se puede clasificar en cinco estadios según el valor de TFG. Luego, se evaluó además el porcentaje de clasificación correcta según las diferentes fórmulas.

En una segunda etapa se recolectaron datos correspondientes al período 2018-2019 ($n = 78$), que se utilizaron como muestra de validación interna. Esto permitió evaluar la performance de las ecuaciones en un conjunto de datos totalmente independiente del utilizado para su desarrollo.

Los resultados muestran que entre las ecuaciones de uso actual, CKD-EPI es la que mejores predicciones produce. Las ecuaciones obtenidas a partir de modelos de cuasiverosimilitud fueron las de mejor desempeño, siendo esto más notable en la muestra de validación interna que en la de desarrollo.

Palabras Clave: filtrado glomerular, validación cruzada, validación interna, predicción, error de clasificación.

¹e-mail: pehuenf@gmail.com, wdouthat@gmail.com, cchiurciu@gmail.com, mlnores@famaf.unc.edu.ar

Patrones espaciales de la desnutrición infantil en Argentina

MELCONIAN, M.S, FERNANDEZ, M.S, TURJANSKI, P., ALTSZYLER, E., LIBMAN, M., PEREZ, A., NUÑEZ, P., CUETO, G.R. ¹

Grupo de Bioestadística Aplicada, DEGE-DC-ICC, FCEN-UBA/CONICET

El retraso en el crecimiento (baja talla para la edad) es considerado el mejor indicador de las desigualdades en salud infantil, ya que refleja un déficit acumulativo del crecimiento lineal durante los primeros años de vida. Afecta el desarrollo físico, cognitivo, social y emocional, y genera consecuencias adversas en el desarrollo educativo y económico durante la vida adulta. El Programa Sumar [2013-actualidad] dependiente del Ministerio de Salud de Argentina, recopila datos antropométricos provenientes de controles de salud de niñas y niños con cobertura exclusiva en el sistema de salud público del país, proporcionando un gran volumen de información que hoy en día se encuentra subanalizada, sobre la población infantil más vulnerable. Con el objetivo de identificar regiones del país con mayor prevalencia de baja talla, se propone estudiar la variación espacial de este indicador en la población de niños y niñas beneficiarios del Programa. Como fuente de datos se utilizaron los últimos registros de talla provenientes de controles de salud de niños y niñas de 0 a 5 años del Programa del período ene/2019 - abr/2020. Para cada uno de los 6.194 centros de salud (efectores) que reportaron controles en este período se calculó la prevalencia de baja talla (PBT) como la cantidad de beneficiarios con baja talla para la edad (según el criterio establecido por la Organización Mundial de la Salud) sobre el total de beneficiarios que realizaron controles de salud, la edad promedio y la proporción de varones. Para analizar los patrones espaciales en la PBT (variable respuesta) se aplicaron modelos geoespaciales con métodos de estimación bayesianos (Bayesian space-model-based geostatistics using integrated nested Laplace approximations), disponibles en la biblioteca R-INLA del software R, utilizando una distribución Binomial. Para el modelado de la estructura espacial se utilizaron ecuaciones estocásticas parcialmente diferenciales, una malla conteniendo las coordenadas de cada uno de los efectores y una matriz de correlación de Matérn. Se incluyeron como variables de control la proporción de varones y la edad promedio de los beneficiarios de cada efector. Como variables de efectos aleatorios se añadieron el efector y el departamento al cual correspondía el mismo. La cantidad de beneficiarios por efector fue variable (media=188, mediana=102; Rango: 1-7.489). En cuanto a lo estimado por el modelo, la mediana para la PBT estimada fue de 8% y los percentiles 5 y 95 fueron de [6,28-10,40] % respectivamente. El modelo ajustado presentó una estructura espacial significativa, (DIC al incorporar el factor espacial: 30.039). La PBT disminuyó al aumentar la edad promedio de los beneficiarios del efector y aumentó con la proporción de varones del mismo. Los valores más altos de PBT se observaron en las provincias del noroeste, noreste, Santiago del Estero y el AMBA, seguidos por la región patagónica, particularmente en la provincia de Neuquén. Modelar la estructura espacial de variables con distribución no normal requiere de métodos complejos existiendo pocos ejemplos de aplicaciones en epidemiología en Argentina. En este sentido, en el presente trabajo el modelo ajustado permitió describir la estructura espacial presente en el indicador de baja talla e identificar áreas a distintas escalas jurisdiccionales con peores situaciones nutricionales. El hecho de ajustar el modelo por edad y sexo (proporción de varones), hizo posible efectuar comparaciones independientemente de la composición de edad y sexo de cada efector. Poder hacer un seguimiento de indicadores de desnutrición a escalas subnacionales y locales es fundamental tanto para reducir las desigualdades a nivel territorial, como para evaluar progresos y áreas que requieren mejoras; además permite implementar políticas sociales que se adecúen a las necesidades particulares de cada caso. De esta manera, este trabajo podrá ser utilizado por tomadores de decisiones como herramienta que permita enfocar posibles intervenciones sanitarias acordes a la situación que se presente, en grupos determinados o espacios geográficos particulares. **Palabras Clave: malnutrición infantil, INLA, análisis espacial.**

¹e-mail: agustinamelco@gmail.com

Evaluación de un índice como marcador de Insulino Resistencia: una aplicación de curvas ROC.

Claudia Molinari 1, Valeria Hirschler 2, Edith Scaiola 3, Cecilia Miorin 3, Zelmira Guntsche MD4, Silvia Lapertosa 2, Claudio Gonzalez 2, Cariño study group¹

1. Facultad de Farmacia y Bioquímica, Dpto Físico-Matemática, Universidad de Buenos Aires

2. Universidad de Buenos Aires 3. Universidad Nacional de La Plata 4. Universidad Nacional de Cuyo

Los métodos considerados como “patrón de oro” para la determinación de insulino resistencia son complejos y de alto costo, especialmente para países sub desarrollados. El índice HOMA es un test derivado de la glucosa y del nivel de insulina que está aceptado y es muy utilizado en la clínica pediátrica. Sin embargo, al necesitar de la determinación de la insulina sérica este índice no resulta fácil de reproducir en cualquier laboratorio. Luego resulta de interés probar alternativas más accesibles para zonas o países de menores recursos. En este contexto surge la propuesta del índice TyG que se basa en el valor de los triglicéridos y de la glucosa, resultando así un procedimiento estandarizado y de mayor disponibilidad. En este trabajo se pone a prueba la efectividad del índice TyG como marcador de Insulino Resistencia (IR) en niños escolarizados de Argentina. Se trabajó con una muestra de 915 niños de edades entre 5 y 13 años. El estudio fue parte de un proyecto multicéntrico que incluye cuatro zonas de la Argentina. Se determinaron medidas antropométricas como también indicadores socioeconómicos de las familias de donde provenían. Se determinaron glucosa, lípidos y se estimó el índice HOMA-IR a partir de la insulina y la glucosa. Antes del inicio del estudio le fue explicado el mismo a cada padre, los cuales dieron su consentimiento por escrito. Para el análisis de datos se realizó una estadística descriptiva exploratoria para la muestra total y para la muestra dividida por género y por edad. Se realizaron test de Student y análisis de varianza para comparar los grupos. Para determinar la performance del índice TyG como marcador de Insulino Resistencia se aplicaron regresiones logísticas y se construyeron curvas ROC para las cuales se analizó el área bajo la curva, la especificidad y sensibilidad. Se obtuvo una alta asociación entre IR y el índice TyG. La performance del índice resultó diferente según se consideren los niños por sexo y edad. Se concluye que el índice es sensible para la población en estudio, aunque en algunos grupos su utilidad resulta limitada.

Palabras Clave: Insulino resistencia, marcadores, curvas Roc

¹e-mail: cmolinari@ffyb.uba.ar