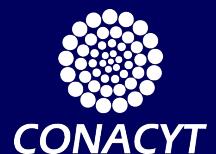




33 F O R O NACIONAL DE ESTADÍSTICA

C O N G R E S O
LATINOAMERICANO
DE SOCIEDADES DE
ESTADÍSTICA 13

Libro de Resúmenes



CONTENIDO

Presentación	3
Directorio Universidad de Guadalajara	4
Directorio AME	5
Organizadores	6
Patrocinadores	7
Actividades de Interés General	8
Cursos Pre-Foro	11
Conferencias Magistrales	16
Cursos Cortos	21
Sesiones Temáticas	28
Ponencias (Contribuciones Libres)	43
Carteles (Contribuciones Libres)	110

PRESENTACIÓN

La Asociación Mexicana de Estadística (AME) tiene el gusto de recibir a la comunidad de profesionales, investigadores y estudiantes de la Estadística en el Foro Nacional (FNE) que en 2018 alcanza su edición 33. Esta ocasión es única y especialmente propicia para el intercambio de experiencias con colegas de otros países de la región latinoamericana en tanto que el FNE se funde, en una sola reunión, con el 13 Congreso Latinoamericano de Sociedades de Estadística (CLATSE).

La institución sede del evento es la Universidad de Guadalajara que amablemente nos abre las instalaciones del Centro Universitario de Ciencias Exactas e Ingeniería.

El Programa Científico ha sido cuidadosamente diseñado por la AME con la colaboración de las Sociedades de Estadística de Argentina, Brasil, Colombia y Chile, y contempla cuatro Conferencias Magistrales ofrecidas por distinguidos expertos, cuatro cursos Pre Foro, cinco Cursos Cortos y seis Sesiones Temáticas, así como 152 Contribuciones Libres agrupadas en 43 Sesiones de Ponencias y una Sesión de Carteles.

La AME agradece al Centro de Investigación en Matemáticas, al Consejo Nacional de Ciencia y Tecnología (proyecto 296890), al Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas de la UNAM, al Instituto Tecnológico Autónomo de México y la Universidad de Guadalajara por su apoyo financiero y logístico para la realización de este evento. Asimismo hacemos patente nuestro agradecimiento a todos nuestros miembros institucionales: Colegio de Posgraduados, Instituto Nacional de Estadística y Geografía, Instituto Nacional para la Evaluación de la Educación, Instituto Tecnológico Autónomo de México, Facultad de Ciencias, Numérica y la Universidad Veracruzana por su constante e invaluable apoyo.

Finalmente, agradecemos a todos y cada uno de los y las colegas que conformaron los comités organizador y científico por su invaluable esfuerzo para que este evento se llevara a cabo exitosamente. Gracias a todos los ponentes por su entusiasta participación en el evento y a todos aquellos que de una u otra forma hicieron posible la realización de este evento.

Dr. Humberto Gutiérrez Pulido
Centro Universitario de Ciencias Exactas e Ingeniería
Jefe del Departamento de Matemáticas

Dr. Ramsés H. Mena Chávez
Asociación Mexicana de Estadística
Presidente

Guadalajara, Jalisco, Septiembre de 2018

DIRECTORIO UNIVERSIDAD DE GUADALAJARA

Dr. Miguel Ángel Navarro Navarro

Rector General

CENTRO UNIVERSITARIO DE CIENCIAS EXACTAS E INGENIERÍA

Dra. Ruth Padilla Muñoz

Rectora del Centro

Dr. Carlos Pelayo Ortíz

Secretario Académico

Mtro. Jaime Gutiérrez Chávez

Secretario Administrativo

Dr. Oscar Blanco Alonso

Director de la División de Ciencias Básicas

Dr. Humberto Gutiérrez Pulido

Jefe del Departamento de Matemáticas

DIRECTORIO AME

Dr. Ramsés H. Mena

Presidente

Dr. Luis Enrique Nieto Barajas

Vicepresidente

Dr. Gabriel Rodríguez Yam

Ex-Presidente

Dra. Isadora Antoniano Villalobos

Dra. Julia Aurora Montano Rivas

Dra. Lizbeth Naranjo Albarán

Dr. Eliud Silva Urrutia

Dr. Asael Fabian Martínez Martínez

Dr. Paulino Pérez Rodríguez

Dra. Lilia Leticia Ramírez Ramírez

Dra. Luz Judith Rodríguez Esparza

Vocales

Dra. Luz Judith Rodríguez Esparza

Tesorera

Luis Enrique Reyes Romero

Secretario de la Junta Directiva

Elida Estrada Barragán

Asistente

ORGANIZADORES

Comité Organizador

Humberto Gutiérrez Pulido, UDG (Presidente)

Leticia Ramírez Ramírez, CIMAT

Comité Científico del FNE

Eduardo Gutiérrez Peña, IIMAS-UNAM

Abelardo Montesinos López, UDG

Lizbeth Naranjo Albarrán, FC-UNAM

Luis Enrique Nieto Barajas, ITAM (Presidente)

Manuel Mendoza, ITAM

Comité Científico del CLATSE

Jorge Luis Bazán, Universidad de São Paulo

Ramón Giraldo, Universidad Nacional de Colombia

Manuel Mendoza, ITAM (Presidente)

Orietta Nicolis, Universidad de Valparaíso

Lila Ricci, Universidad Nacional del Mar de Plata

PATROCINADORES



ACTIVIDADES DE INTERÉS GENERAL

Registro de Participantes

Cursos Pre-Foro

Lunes 1 de octubre

Auditorios Antonio Alatorre y Antonio Rodríguez

FNE/CLATSE

Del 2 al 5 de octubre, 8:00 hrs.

A un costado del Módulo Y

Cursos Pre-Foro

Lunes 1 de octubre

Auditorios Antonio Alatorre y Antonio Rodríguez

Ceremonia de Inauguración

Martes 2 de octubre

9:00 a 10:00 hrs

Auditorio Módulo Y

Sesión de Carteles y Brindis de Bienvenida

Martes 2 de octubre

19:00 a 21:00 hrs

Auditorio E. Díaz de León

Sesión Especial del Premio FAO

Jueves 4 de octubre

12:30 a 14:00 hrs

Auditorio Matute Remus

Asamblea General Ordinaria de la AME/CLASE

Jueves 4 de octubre

19:00 a 23:00 hrs

Auditorio Módulo Y

Ceremonia de Clausura

Viernes 5 de octubre

19:00 a 19:30 hrs

Auditorio Módulo Y

PROGRAMA GENERAL

Lunes 1	Martes 2	Miércoles 3	Jueves 4	Viernes 5
CURSOS PRE-Foro 8:00 hrs (Receso para café de 10:30 a 11:00)	CEREMONIA DE INAUGURACIÓN 9:00 hrs	CURSOS CORTOS 8:00 hrs	CURSOS CORTOS 8:00 hrs	CURSOS CORTOS 8:00 hrs
	CONFERENCIA MAGISTRAL 10:00 hrs	CONFERENCIA MAGISTRAL 10:00 hrs	CONFERENCIA MAGISTRAL 10:00 hrs	CONFERENCIA MAGISTRAL 10:00 hrs
	CONTRIBUCIONES LIBRES 11:00 hrs	CONTRIBUCIONES LIBRES 11:00 hrs	CONTRIBUCIONES LIBRES 11:00 hrs	CONTRIBUCIONES LIBRES 11:00 hrs
	Receso para café (12:00 hrs)	Receso para café (12:00 hrs)	Receso para café (12:00 hrs)	Receso para café (12:00 hrs)
	SESIONES TEMÁTICAS 12:30 hrs	SESIONES TEMÁTICAS 12:30 hrs	SESIONES TEMÁTICAS Y SESIÓN PREMIO FAO 12:30 hrs	CONTRIBUCIONES LIBRES 12:30 hrs
Receso para comida 13:30 hrs	Receso para comida 14:00 hrs	Receso para comida 14:00 hrs	Receso para comida 14:00 hrs	Receso para comida 14:00 hrs
	CONTRIBUCIONES LIBRES 16:00 hrs	CONTRIBUCIONES LIBRES 16:00 hrs	CONTRIBUCIONES LIBRES 16:00 hrs	CONTRIBUCIONES LIBRES 16:00 hrs
CURSOS PRE-Foro 15:30 hrs (Receso para café de 18:00 a 18:30)	Receso para café (17:00 hrs)	Receso para café (17:00 hrs)	Receso para café (17:00 hrs)	Receso para café (17:00 hrs)
	CONTRIBUCIONES LIBRES 17:30 hrs	CONTRIBUCIONES LIBRES 17:30 hrs	CONTRIBUCIONES LIBRES 17:30 hrs	CONTRIBUCIONES LIBRES 17:30 hrs
	BRINDIS Y SESIÓN DE CARTELES 19:00 hrs	ASAMBLEA GENERAL AMÉ/CLASE 19:00 hrs	CEREMONIA DE CLAUSURA 19:00 hrs	



CURSOS PRE-FORO

LUNES 1 DE OCTUBRE
8:00 A 13:30 HRS

CURSOS PRE-FORO
AUDITORIO ANTONIO ALATORRE

— I —

Variational Bayes and beyond: Bayesian inference for big data

Impartido por

Tamara Broderick

Massachusetts Institute of Technology, USA

Moderador: **Alexandra M. Schmidt** (Mc Gill University)

RESUMEN. Bayesian methods exhibit a number of desirable properties for modern data analysis—including (1) coherent quantification of uncertainty, (2) a modular modeling framework able to capture complex phenomena, (3) the ability to incorporate prior information from an expert source, and (4) interpretability. In practice, though, Bayesian inference necessitates approximation of a high-dimensional integral, and some traditional algorithms for this purpose can be slow—notably at data scales of current interest. The tutorial will cover modern tools for fast, approximate Bayesian inference at scale. One increasingly popular framework is provided by “variational Bayes” (VB), which formulates Bayesian inference as an optimization problem. We will examine key benefits and pitfalls of using VB in practice, with a focus on the widespread “mean-field variational Bayes” (MFVB) subtype. We will highlight properties that anyone working with VB, from the data analyst to the theoretician, should be aware of. We will cover modern corrections to VB for the purposes of uncertainty and robustness quantification. In addition to VB, we will cover recent data summarization techniques for scalable Bayesian inference that come equipped with finite-data theoretical guarantees on quality. We will motivate our exploration throughout with practical data analysis examples and point to a number of open problems in the field.

LUNES 1 DE OCTUBRE
8:00 A 13:30 HRS

CURSOS PRE-FORO
AUDITORIO ANTONIO RODRÍGUEZ

— II —

Machine Learning

Impartido por

Elmer Garduño

Google Inc., USA

Moderador: **Luis Enrique Nieto Barajas** (Instituto Tecnológico Autónomo de México)

RESUMEN. Learn how to build Machine Learning applications with Tensorflow, this hands-on workshop will walk you through how to use Tensorflow to build classification, regression and recommendation models using Tensorflow. We will also talk about building and deploying production models and reusing ML models with Tensorflow Hub.

El curso será dictado en español.

LUNES 1 DE OCTUBRE
15:30 A 21:00 HRS

CURSOS PRE-FORO
AUDITORIO ANTONIO RODRÍGUEZ

— III —

Bayesian computing with INLA

Impartido por

Haavard Rue

King Abdullah University of Science and Technology, SAU

Moderador: **Ramsés Mena Chávez** (Universidad Nacional Autónoma de México)

RESUMEN. In this course, we will discuss approximate Bayesian inference for a class of models named latent Gaussian models (LGM). LGM's are perhaps the most commonly used class of models in statistical applications. It includes, among others, most of (generalized) linear models, (generalized) additive models, smoothing spline models, state space models, semiparametric regression, spatial and spatio-temporal models, log-Gaussian Cox processes and geostatistical and geoadditive models. The concept of LGM is intended for the modelling stage, but turns out to be extremely useful when doing inference as we can treat models listed above in a unified way and using the *same* algorithm and software tool. Our approach to (approximate) Bayesian inference, is to use integrated nested Laplace approximations (INLA). Using this tool, we can directly compute very accurate approximations to the posterior marginals. The main benefit of these approximations is computational: where Markov chain Monte Carlo algorithms need hours or days to run, our approximations provide more precise estimates in seconds or minutes. Another advantage with our approach is its generality, which makes it possible to perform Bayesian analysis in an automatic, streamlined way, and to compute model comparison criteria and various predictive measures so that models can be compared and the model under study can be challenged. In this course, we will introduce the class of latent Gaussian models, describe the “big picture” of the INLA algorithm and introduce the R-INLA package. We will focus on applied aspect and the use of the package illustrated on several examples.

Este curso será impartido conjuntamente con Daniela Castro-Camilo.

LUNES 1 DE OCTUBRE
15:30 A 21:00 HRS

CURSOS PRE-FORO
AUDITORIO ANTONIO ALATORRE

— IV —

Statistical and psychometric intricacies of educational survey assessments

Impartido por
Andreas Oranje
Educational Testing Service, USA

Moderador: **Luis Enrique Nieto Barajas** (Instituto Tecnológico Autónomo de México)

RESUMEN. During this course, we will briefly introduce the core goals of educational survey assessments (also known as group score assessments or large-scale assessments) and the most common designs to meet those goals. From there, we will discuss the core psychometric and statistical principles, including the use of item response theory and latent regression analysis to develop key group statistics of interest. The remainder of the course will focus on statistical topics including

1. typical item and population sampling designs
2. statistical inference (including the three main components of variance and how they are computed, estimation of degrees of freedom, multiple comparisons, and various statistical rules that are often applied).

We will end with a brief overview of where these assessments are heading in terms of digitization, the use of behavioral and process data, and using adaptive approaches to test design and administration. At the end of this course, we hope that participants can more quickly read psychometric and statistical papers about group score assessment and possibly gain an interest in working on and developing these types of assessments. Besides a set of slides, we will provide access to a new summary article about NAEP statistical and psychometric research (Oranje and Kolstad, in press, Special Issue of Journal of Educational and Behavioral Statistics) as well as a bibliography with easily accessible/downloadable papers that address the aforementioned topics in more detail.

CONFERENCIAS MAGISTRALES

MARTES 2 DE OCTUBRE
10:00 A 11:00 HRS

CONFERENCIAS MAGISTRALES
AUDITORIO MÓDULO Y

— I —

Non-Gaussian spatial and spatio-temporal processes

Impartido por

Alexandra M. Schmidt

McGill University

Moderador: **Luis Enrique Nieto Barajas** (Instituto Tecnológico Autónomo de México)

RESUMEN. In the analysis of most spatial and spatio-temporal processes in environmental studies, observations present skewed distributions, with a heavy right or left tail. Usually, a single transformation of the data is used to approximate normality, and stationary Gaussian processes are assumed to model the transformed data. Spatial interpolation and/or temporal prediction are routinely performed by transforming the predictions back to the original scale. The choice of a distribution for the data is key for spatial interpolation and temporal prediction. In this talk, I will start discussing the advantages and disadvantages of using a single transformation to model such processes. Then I will discuss some recent advances in the modeling of non-Gaussian spatial and spatio-temporal processes.

MIÉRCOLES 3 DE OCTUBRE
10:00 A 11:00 HRS

CONFERENCIAS MAGISTRALES
AUDITORIO MÓDULO Y

— II —

Bayesian quantile regression for discrete observations

Impartido por

Haavard Rue

King Abdullah University of Science and Technology, Saudi Arabia

Moderador: **Mauricio Castro Cepero** (Pontificia Universidad Católica de Chile)

RESUMEN. Quantile regression, i.e. modeling conditional quantiles of some covariates and other effects through the linear predictor, has typically been carried out exploiting the asymmetric Laplace distribution (ALD) as a working “likelihood”. In the Bayesian framework, this is highly questionable as the posterior variance is affected by the artificial ALD “likelihood”. With continuous responses, we can reparameterize the likelihood in terms of a alpha-quantile, and let the alpha-quantile depend on the linear predictor. We can then do model based quantile regression with little effort using the R-INLA package (www.r-inla.org) doing approximate Bayesian inference for latent gaussian models, and trust the quantile regression posterior in the same way as when doing parametric mean regression. For discrete variables, like Poisson and (negative) Binomial, there is no continuous relationship between quantiles and distribution’s parameters, hence model based quantile regression seems no longer possible. In this talk I will discuss how to resolve this issue, so that we can do model based quantile regression also for discrete responses. I will present some examples that also demonstrate how the parametric approach almost resolves the quantile crossing problem.

JUEVES 4 DE OCTUBRE
10:00 A 11:00 HRS

CONFERENCIAS MAGISTRALES
AUDITORIO MÓDULO Y

— III —

Have you cake and eat it: Flexibility vs. interpretability in Data Science

Impartido por

Abel Rodriguez

University of California, Santa Cruz

Moderador: **Manuel Mendoza Ramírez** (Instituto Tecnológico Autónomo de México)

RESUMEN. Over the last 10 years, Deep Learning (DL) methods (which are based on flexible models build by the repeated composition of non-linear kernels) have shown success in tasks as diverse as image recognition and automatic machine translation. Although this success has made DL extremely popular, the methods present important practical drawbacks. Of these, the inability of DL to provide “human-friendly” explanations is one of the most serious. Consider, for example, so-called autoencoders, which aim to learn a sparse, low-dimensional representation for a set of observations. Unlike standard statistical methods such as principal component or factor analysis, the low-dimensional features constructed by an autoencoder typically have no straightforward interpretation in terms of the original application. While this is usually not a problem if the features will be used to construct a predictive model, the lack of interpretability is a serious issue in many social science applications where the main interest lies precisely on the latent features. This talk will provide an example of how carefully-constructed statistical methods can play a key role in the derivation of data science methods that are both flexible and interpretable. In particular, we will introduce a new class of embedding models for matrix-valued binary data, and discuss a particular instantiation of the class that is motivated by practical issues associated the analysis of legislative voting data in the United States and the United Kingdom. The resulting model is a generalization of the class of spatial voting model from political science that allows the preference space to be any arbitrary Riemannian manifolds. The features of the model are illustrated using a number of datasets of voting records from the US House of Representatives.

VIERNES 5 DE OCTUBRE
10:00 A 11:00 HRS

CONFERENCIAS MAGISTRALES
AUDITORIO MÓDULO Y

— IV —

Statistical innovation in Brazil

Impartido por

Francisco Louzada

ICMC-USP

Moderador: **Ramsés Mena Chávez** (Universidad Nacional Autónoma de México)

RESUMEN. Statistics has been consolidated as an activity to promote scientific and technological progress in Brazil, contributing to the productive sector in order to increase its added value through the development of specific methodologies and products. In this context, we have observed a great insertion of methodologies and statistical products in innovation processes, promoting interaction with professionals from the public and productive sectors, as well as with the population in general, which efficiently directs the interlocution between academia and industry. This conference presents the main actions and activities of statistical innovation that we have been developing within the Center for Mathematics and Statistics Applied to Industry (CeMEAI), in the sense of approximating academia, the productive sector and the community in general. The social, cultural and financial impacts of our actions and activities are exemplified through real success examples. This includes reliability modeling for oil well construction equipment, classification modeling for fraud detection in financial transactions, wood quality automatic recognition modeling, solid waste management strategies, electro-encephalography modeling, breast cancer modeling, neglected tropical diseases modeling, communication modeling for mobile phones and autonomous unmanned aerial vehicles, and sport data modeling. Challenges and future trends for innovation and transfer of technology in Statistics are also addressed.

CURSOS CORTOS

— I —

Estadística espacial

Impartido por

Ronny Vallejos

Universidad Técnica Federico Santa María, Chile

Moderador: **Gabriel Rodríguez Yam** (Universidad Autónoma Chapingo)

RESUMEN. El objetivo principal de este curso es introducir algunos métodos que se han desarrollado en la literatura en los últimos 25 años, respecto a la cuantificación de la correlación entre dos procesos espaciales. Primero, se discutirán los conceptos básicos que se usan en Estadística Espacial como los procesos intrínsecamente estacionarios, el variograma y el variograma cruzado. Entonces, la metodología sería abordada en dos grupos de técnicas. El primer grupo aglomera la problemática desde una perspectiva de contraste de hipótesis acerca del parámetro de la correlación entre dos procesos Gaussianos. El segundo grupo de técnicas consiste en variantes del coeficiente de correlación de Pearson. En ambos casos, se presentarán ilustraciones con datos reales que consisten en realizaciones de procesos espaciales en grillas regulares (imágenes) y grillas no regulares. La mayoría de las aplicaciones que se presentarán serán abordadas usando las técnicas implementadas en los paquetes de R SpatialPack y GeoModels. El curso será autocontenido orientado a una audiencia general, pero al mismo tiempo se expondrá en la parte final algunos tópicos que son materia de investigación reciente y algunos problemas abiertos.

CONTENIDO

- Introducción al problema de comparación entre dos procesos espaciales desde el punto de vista de la correlación y algunos ejemplos. Introducción al análisis espacial, procesos espaciales, estacionariedad intrínseca, el variograma y su estimación, Kriging, el variograma cruzado y análisis de imágenes.
- Test de asociación espacial. El test t , el test t modificado, el tamaño muestral efectivo, un test t basado en permutaciones, extensión del test t para cuantificar la asociación entre un proceso y un vector de procesos. Aplicaciones en R en un contexto de datos forestales y de contaminación del suelo. Test del coeficiente de correlación de un proceso normal bivariado. Distribución asintótica asociada al test. Desempeño del test cuando hay desviaciones de la normalidad. Extensiones a procesos espaciales con distribución t -student. Experimentos computacionales basados en el paquete de R GeoModels.
- El coeficiente de codispersión y sus propiedades. Distribución asintótica y test de hipótesis. Representación de la codispersión en el plano y el mapa de codispersión. Índices de similaridad entre imágenes. Aplicaciones en R.

BIBLIOGRAFÍA

1. Clifford, P., Richardson, S., Hémon, D. (1989). Assessing the significance of the correlation between two spatial processes. *Biometrics* 45, 123-134.
2. Bevilacqua, M., Vallejos, R., Velandia, D. (2015). Assessing the significance of the correlation between the components of a bivariate Gaussian random field. *Environmetrics* 26, 545-556.
3. Ojeda, S., Britos, G., Vallejos, R. (2018). An image quality index based on coefficients of spatial association with an application to image fusion. *Spatial Statistics* 23, 1-16.
4. Rukhin, A. Vallejos, R. (2008). Codispersion coefficient for spatial and temporal series. *Statistics and Probability Letters* 78, 1290-1300.
5. Vallejos, R., Osorio, F., (2014). Effective sample size for spatial process models. *Spatial Statistics* 9, 66-92.
6. Vallejos, R., Mancilla, D., Acosta, J. (2016). Image similarity assessment based on measures of spatial association. *Journal of Mathematical Imaging and Vision* 56, 77-98.

3–5 OCTUBRE
8:00 A 9:00 HRS

CURSOS CORTOS
AUDITORIO ANTONIO ALATORRE

— II —

Computational methods for Bayesian inference

Impartido por

Hedibert Lopes

Insper, Brasil

Moderador: **Carlos Erwin Rodríguez Hernández-Vela** (Universidad Autónoma Metropolitana,
Iztapalapa)

CONTENTS

- Bayesian inference via Gibbs, Metropolis and other MCMC schemes
- Sequential Monte Carlo for Bayesian sequential learning
- Bayesian inference and computation in econometrics

All three lectures will start with simple, toy examples to introduce concepts and illustrate its implementation and will finish with more serious, complex case studies based on my own research over the years.

3–5 OCTUBRE
9:00 A 10:00 HRS

CURSOS CORTOS
AUDITORIO ANTONIO ALATORRE

— III —

Estadística ambiental

Impartido por

Bruno Sansó

University of Santa Cruz, EUA

Moderador: **Lizbeth Naranjo Albarrán** (Universidad Nacional Autónoma de México)

CONTENIDO

- En la primera clase cubriré los métodos para inferencia y predicción para procesos espaciales usando en enfoque Bayesiano
- En la segunda hablaré de modelos de rango fijo: procesos de convolución y procesos Gaussianos predictivos
- En la tercera clase hablaré de modelos espacio-temporales

3–5 OCTUBRE
9:00 A 10:00 HRS

CURSOS CORTOS
AUDITORIO ANTONIO RODRÍGUEZ

— IV —

Retos en la enseñanza de la Estadística: Nuevos escenarios en grado, master y doctorado

Impartido por

María Purificación Galindo

Universidad de Salamanca, España

Moderador: **Gabriel Rodríguez Yam** (Universidad Autónoma Chapingo)

RESUMEN. Por anunciarse

— V —

Bases estadísticas del aprendizaje de máquina con STATA

Impartido por

Miguel Ángel Cruz

MultiON Consulting, México

Moderador: **Luz Judith Rodriguez Esparza** (Universidad Autónoma Chapingo)

RESUMEN. El machine learning es un conjunto de técnicas estadísticas que permiten crear algoritmos para el análisis de datos, en donde un modelo “es entrenado” mediante datos existentes, y que tiene como fin último utilizar dicho modelo “entrenado” para predecir objetivos o nuevos datos. Este enfoque de prueba y error denota dos tipos de problemas sobre el aprendizaje del modelo: supervisado y no supervisado; mismos que se estudian y solucionan con tres grandes bloques analíticos de la estadística: agrupamiento de datos, clasificación de datos y análisis de regresión. Este curso corto pretende introducir a los participantes en estos tres bloques de análisis multivariante, y en sus distintos métodos, con la finalidad de que las bases estadísticas del machine learning queden claras cuando el participante se introduzca en el análisis de los algoritmos desarrollados por este enfoque.

CONTENIDO

- Introducción
- Análisis clúster con STATA
- Análisis de clasificación con STATA
- Análisis de regresión con STATA

SESIONES TEMÁTICAS

— I —

An overview of phase-type distributions: Statistical inference and applications

Organizador: **Jorge González** (Pontificia Universidad Católica de Chile)

Moderador: **Gabriel Núñez Antonio** (Universidad Autónoma Metropolitana, unidad Iztapalapa)

1.1 *Bayesian inference in phase-type distributions*

Luis Gutiérrez (Pontificia Universidad Católica de Chile)

RESUMEN. Phase-type distributions have been studied mainly in the applied probability literature. This class of distributions is dense in the space of probability measures defined in the positive real numbers. The above property, together with many other features, which they inherit from the underlying Markov jump process, make the class appealing in a variety of statistical problems. Although Phase-type distributions enjoy appealing features, the inference is still challenging due to the latent nature of the Markov jump process. In this talk, we will discuss Bayesian approaches of inference with a special emphasis on identifiability issues, computational complexity, exibility and scalability. Posterior algorithms for finite mixtures of Phase-type distributions will be presented. Additionally, a particular application in the context of renewal theory will be developed.

1.2 *Fitting phase-type scale mixtures to heavy-tailed risks*

Leonardo Rojas Nandayapa (IIMAS-UNAM)

RESUMEN. We consider the fitting of heavy tailed risk distributions with a special attention to distributions with a non-standard shape in the “body” of the distribution. To this end we consider a dense class of heavy tailed distributions introduced recently, employing an EM algorithm for the maximum likelihood estimates of its parameters. We present methods for fitting to observed data, histograms, censored data, as well as to theoretical distributions. Numerical examples are provided with simulated data and a benchmark reinsurance dataset. We empirically demonstrate that our model can provide excellent fits to heavy-tailed data/distributions with minimal assumptions.

1.3 *Some applications of phase-type distributions*

Jorge González (Pontificia Universidad Católica de Chile)

RESUMEN. Phase-Type distributions (PH) describe absorption times in Markov processes with a finite number of transient states and one absorbing state. A remarkable property that makes these distributions versatile is that they are dense and can be used to approximate any kind of distribution on the positive real numbers. Applications of PH distributions can be encountered in various areas such as the analysis of manufacturing systems, risk models, performance and dependability analysis of computer and communication systems, analysis of healthcare systems, infection models, renewal

theory, and approximation of Heavy-tailed distributions, among others. In this talk, we will show some examples where PH distributions have been successfully applied. The potential use of these type of distributions in other fields will be also discussed.

MARTES 2 DE OCTUBRE
12:30 A 14:00 HRS

SESIONES TEMÁTICAS
AUDITORIO MÓDULO Y

— II —

Grafos probabilísticos

Organizador: **Juan Carlos Martínez-Ovando** (ITAM, México)

Moderador: **Isadora Antoniano Villalobos** (Bocconi University)

2.1 *Applications of graphical models in energy markets*

Enrique Sucar (INAOE, México)

RESUMEN. Por anunciarse

2.2 *Neural Networks, Cognitive Systems and causality*

Nihat Ay (Santa Fe Institute, EUA)

RESUMEN. Por anunciarse

2.3 *Automated financial fraud and collusion detection*

Juan Carlos Martínez-Ovando (ITAM, México)

RESUMEN. Por anunciarse

— III —

Análisis estadístico para datos complejos

Organizador: **Mauricio Castro** (Pontificia Universidad Católica de Chile)

Moderador: **Ana Belém Nieto Librero** (Universidad de Salamanca)

3.1 *A Bayesian analysis of the matching problem*

Ignacio Vidal (Universidad de Talca, Chile)

RESUMEN. Reliability of measurements is a fundamental prerequisite in any research that we want to develop in the field of experimental sciences, like for example social science, statistical process control, physic, metrology, biological and medical sciences, etcetera. Multiple statistical strategies exist in the literature to assess the agreement between two or more measurement methods (instruments or observers), especially for interval or ratio data. There is a less variety of statistical methods to assess the agreement between observations with nominal or ordinal levels of measurement, a few examples are Kendall's coefficient, kappa's coefficient, McNemar test, chi square test, etc. A less common experiment to assess the agreement between categories of nominal level versus gold standard classifications is the one that gives rise to the matching problem. The matching problem is known since the beginning of the eighteenth century and a Bayesian analysis motivated by this problem can be seen in Diaconis and Holmes (2002). In this work we present the Bayesian analysis of an experiment that also leads to the matching problem. However, we have considered dependence on different classifications. Conjugate prior distributions for the likelihood function have been discussed. The frequentist analysis has a particular Bayesian interpretation under a uniform prior distribution. Finally, by Bayes risk minimization an optimum sample size is obtained.

3.2 *A Bayesian nonparametric model for predicting pregnancy outcomes using longitudinal profiles*

Rolando de la Cruz (Universidad Adolfo Ibañez, Chile)

RESUMEN. Across several medical fields, developing an approach for disease classification is an important challenge. The usual procedure is to fit a model for the longitudinal response in the healthy population, a different model for the longitudinal response in disease population, and then apply the Bayes' theorem to obtain disease probabilities given the responses. Unfortunately, when substantial heterogeneity exists within each population, this type of Bayes classification may perform poorly. In this work, we develop a new approach by fitting a Bayesian nonparametric model for the joint outcome of disease status and longitudinal response, and then use clustering induced by the Dirichlet process in our model to increase the flexibility of the method, allowing for multiple subpopulations of healthy, diseased, and possibly mixed membership. In addition, we introduce an MCMC sampling scheme that facilitates the assessment of the inference and prediction capabilities of our model. Finally, we demonstrate the method by predicting pregnancy outcomes using longitudinal profiles on

the β -HCG hormone levels in a sample of Chilean women being treated with assisted reproductive therapy.

3.3 *Automated learning of t factor analysis models with complete and incomplete data*

Mauricio Castro (Pontificia Universidad Católica de Chile)

RESUMEN. The t factor analysis (tFA) model has been shown a promising tool for robust reduction of high-dimensional data in the presence of heavy-tailed noises. When determining the number of factors of the tFA model, a two-stage procedure is commonly performed in which parameter estimation is carried out for a number of candidate models, and then the best model is chosen according to certain penalized likelihood indices such as the Bayesian information criterion. However, the computation burden of such a procedure could be extremely high to achieve the optimal performance, particularly for extensively large datasets. In this paper, we are devoted to developing a novel automated learning method in which parameter estimation and model selection are seamlessly integrated in a one-stage algorithm. This new scheme is called the automated tFA (AtFA) algorithm, and it is also workable when missing values exist in the data. In addition, we derive the Fisher information matrix to approximate the asymptotic covariance matrix associated with the ML estimators of tFA models. Experiments on real and simulated datasets reveal that the AtFA algorithm not only provides identical fitting results as compared to traditional two-stage procedures but also runs much faster, especially when data contain missing values.

— IV —

Regression models with bounded inflated distributions

Organizador: **Cristian Bayes** (Pontificia Universidad Católica del Perú)

Moderador: **Luis Alberto Firinguetti Limone** (Universidad del Bío-Bío)

4.1 A beta-inflated mean regression model with mixed effects for fractional response variables

Luis Valdivieso (Pontificia Universidad Católica del Perú)

RESUMEN. In this work we propose a new mixed-effects regression model for fractional bounded response variables. Our model allows us to incorporate covariates directly to the expected value, so we can quantify exactly the influence of these covariates in the mean of the variable of interest rather than to the conditional mean. Estimation is carried out from a Bayesian perspective. Due to the complexity of the augmented posterior distribution, we use a Hamiltonian Monte Carlo algorithm, the No-U-Turn sampler, implemented using the Stan software. A simulation study was performed showing that our model has a better performance than other traditional longitudinal models for bounded variables. Finally, we applied our beta-inflated mean mixed-effects regression model to real data which consists of utilization of credit lines in the peruvian financial system.

4.2 Bayesian spatial inflated beta regression model for assessment of reading level in school districts of Perú

Zaida Quiroz (Pontificia Universidad Católica del Perú)

RESUMEN. The Peruvian Ministry of Education organizes from time to time Student Census Evaluations in order to track the national assessment of educational progress and allocate efforts strategically. One relevant variable in this analysis is the reading level and particularly the proportion of high schools that achieves the highest or third reading level (3RL) at each district. If we aim to determine which factors affect to this proportion we could perform a regression analysis. However, such a proportion turns out to be occasionally 0, especially in rural areas commonly associated with low-level reading skills. In addition, the 3RL district proportions may be spatially referenced, that is, influenced by its neighbors. To overcome these problems we propose the use a beta inflated mean regression analysis which allow us to not only manage the 0 values and introduce a spatial effect, but also to incorporate covariates directly to the expected 3RL district proportions. The parameter estimation in the proposed model is performed with Bayesian techniques using the Hamiltonian Monte Carlo algorithm implemented in the Stan software.

4.3 *Por anunciarse*

Giancarlo Sal y Rosas (Pontificia Universidad Católica del Perú)

RESUMEN. Por anunciarse

— V —

Innovation in risk modeling

Organizador: **Vera Tomazella** (Universidade Federal de São Carlos, Brasil)

Moderador: **Rodrigo dos Santos Targino** (Fundação Getulio Vargas)

5.1 *Efficient closed-form MAP estimators for embedded systems*

Francisco Louzada Neto (Universidade de São Paulo, Brasil)

RESUMEN. Embedded systems are commonly used in communication engineering. Particularly, they can consist of an electronic system inside a microcontroller, which can be programmed to maintain communication between a transmitting antenna and mobile antennas, which are operating at the same frequency. In this context, from the statistical point of view, closed-form estimators are needed, since they are embedded in mobile devices and need to be sequentially recalculated at real time. In this talk, we discuss the proposition a maximum a posteriori (MAP) estimator, which has a simple closed-form expression. In principle, we focus on the Nakagami distribution, which plays an essential role in communication engineering problems, particularly to model fading of radio signals. In a second phase of the presentation, we show that the obtained results can be extended to other survival probability distributions, such as the gamma and generalized gamma ones. Numerical results reveal that the MAP estimator outperforms the existing estimators and produces almost unbiased estimates even for small sample sizes. This work is co-authored by Pedro Luiz Ramos, Eduardo Ramos, and Dipak Dey.

5.2 *Zero-inflated survival models in health*

Gleici da Silva Castro Perdoná (FMRP-USP, Brasil)

RESUMEN. A characteristic of the survival models is to consider the variable ($T > 0$) time greater than zero. Eventually, in survival studies, times equal to zero may be part of the nature of the event. For statistical analysis, individuals with times equal to zeros are withdrawn. However, such times may characterize the population differently, such as when the failure time of a kidney transplant is equal to zero, can modify the overall survival of transplant failure times or for example in can modify the survival of labor failures times, because for the fetal death the time is not recorded. In this paper, we discuss several cases in the area of health, where this phenomenon is possible and we present a survival model that includes this phenomenon (times equal to zero), generating what we will call a zeros inflation. We discuss the results by comparing the results of models that do not include this property. In addition, the modeling considers another proportion of individuals who may have their times censored (incomplete) by some type of intervention. Procedures for estimating parameters are presented, as well as interval estimation using asymptotic methodology, bootstrap intervals, hypothesis testing procedure for model parameters and model selection criteria.

5.3 Repairable system subject to competing risks: objective bayesian estimation method

Vera Lucia Damasceno Tomazella (Universidade Federal de São Carlos, Brasil)

RESUMEN. In this work, objective Bayesian methods are proposed to analyze the minimal repair framework for competing risks. We analyzed the recurrence of failures in a reparable system subject to a variety of failure modes and with a minimal repair type of corrective maintenance assumption. The intensity function for each of modes is described by a power law process. The estimation of the parameters of the model is done using objective Bayesian analysis methodology. Under orthogonality of the parameters we obtain a single prior distribution for any parameter of interest of model. We derive other non-informative prior using formal rules, such as Jeffreys prior to show that are remarkably identical. The posterior reference distribution follows a product of independent gamma distributions. The methodology described is applied to a real data set.

— VI —

Bayesian analysis of survey data and machine learning for nested structures

Organizador: **Luis G. León** (University of Texas Health Science Center at Houston, School of Public Health, USA)

Moderador: **Eduardo Gutiérrez Peña** (Universidad Nacional Autónoma de México)

6.1 *A Bayesian, information-theoretic approach to optimal sampling*

Andrew Womack (Indiana University at Bloomington, USA)

RESUMEN. From the Bayesian perspective, ignorable sampling provides no information that affects the posterior distribution or predictions except through the observed values for the sampled units. Without the imposition of additional structure or relying on the frequentist paradigm, it is not clear what can be said to distinguish sampling schemes from a Bayesian perspective. Assuming a known, fixed sample size and given an inferential quantity of interest, we formulate an information theoretical framework that provides an optimal sampling design. We show that the derived optimal designs are unique and discuss some properties of the designs that are obtained. In particular, optimal sampling in this framework is strongly tied to the inferential target and assumptions about the data generating process.

This is joint work with Daniel Manrique-Vallier.

6.2 *Extensions of machine learning methods for classification of objects based on measurements of embedded observations within each object*

Jose Miguel Yamal (University of Texas Health Science Center at Houston, School of Public Health, USA)

RESUMEN. We consider here an optical technology proposed as an accurate and low-cost alternatives for detection of cervical intraepithelial neoplasia (CIN): quantitative cytology. We discuss the development of classification algorithms for the discrimination between high-grade and low-grade pre-cancer. Classification using quantitative cytology involves classifying a patient based on measurements on a collection of a random number of their cells, e.g., classifying a macro-level object based on measurements of embedded (micro-level) observations within each object. Classification problems with this hierarchical, nested structure have not received the same statistical understanding as the general classification problem that does not account for correlations between measurements within the macro-level object. We propose extensions of some machine and statistical learning models to this structure of data and show, using simulations and real data, the benefit of using these methods over standard machine learning methods with macro-level summary features. These methods have broader applications in classification of hierarchical data including hospitals using

electronic medical records, schools using student measurements, and more. Quantitative cytology shows promise for screening for cervical cancer.

6.3 Bayesian Estimation Under Informative Sampling

Luis G. León (University of Texas Health Science Center at Houston, School of Public Health, USA)

RESUMEN. Bayesian estimation is increasingly popular for performing model based inference to support policymaking. These data are often collected from surveys under informative sampling designs (as opposed to under simple random sampling) where subject inclusion probabilities are designed to be correlated with the response variable of interest. Sampling weights constructed from marginal inclusion probabilities are typically used to form an exponentiated pseudo likelihood that adjusts the population likelihood for estimation. We propose an alternative adjustment based on a Bayes rule construction that simultaneously performs weight smoothing and estimates the population model parameters in a fully Bayesian construction. We compare performances between the two approaches, pseudolikelihood vs Bayesian, on synthetic data, which reveals that our fully Bayesian approach better estimates posterior uncertainty. We demonstrate our method on an application concerning the National Health and Nutrition Examination Survey exploring the relationship between caffeine consumption and systolic blood pressure.

This is joint work with: Terrance D. Savitsky from the US Bureau of Labor Statistics.

— VII —

Sesión Especial del Premio FAO

Moderador: **Gabriel Rodríguez Yam** (Universidad Autónoma Chapingo)

7.1 *Inferencia estadística para el proceso de Poisson modulado*

Roberto Cabal López (Universidad Autónoma de Yucatán)

RESUMEN. En esta tesis se estudia al proceso de Poisson modulado (MMPP), un proceso estocástico que es una generalización del proceso de Poisson homogéneo. El MMPP se entiende como un proceso de Poisson en el cual la tasa de ocurrencia de eventos está gobernada por una cadena de Markov a tiempo continuo irreducible y con espacio de estados finito. Es decir, cada cambio de estado de la cadena de Markov (subyacente) provoca que la intensidad de llegadas en un sistema sea mayor o menor. El objetivo de esta tesis es estudiar una metodología ya existente para la estimación de los parámetros del MMPP. Esta metodología consiste en hallar los estimadores de máxima verosimilitud de los parámetros de la cadena de Markov subyacente y de las posibles tasas de ocurrencias de eventos en el sistema. En esta tesis, una muestra se entiende como una realización del proceso de llegadas, i.e., una trayectoria del proceso. Debido a que por lo general no es posible observar la cadena de Markov subyacente, esta metodología introduce la noción de datos no observados, y utiliza al algoritmo EM para hallar de manera iterativa los estimadores de máxima verosimilitud. Adicionalmente, se utiliza al algoritmo de agrupamiento en k medias para establecer los valores iniciales del algoritmo. Para cumplir con este objetivo, en esta tesis se presenta una breve introducción a las cadenas de Markov a tiempo continuo y al MMPP. Además, se realiza un análisis trayectorial de este último con el objetivo de verificar si una serie de datos podría provenir de este proceso. Y adicionalmente se presenta una prueba de hipótesis para descartar que una muestra provenga de un proceso de Poisson homogéneo. Por último, se aplica la metodología presentada para estimar los parámetros de un MMPP cuando éste es adaptado a un problema de tráfico web.

7.2 *Simulación y estimación en tres procesos de difusión fundamentales: Ornstein-Uhlenbeck, Cox-Ingersoll-Ross y Wright-Fisher*

Eduardo Selim Martínez Mayorga (Facultad de Ciencias, UNAM)

RESUMEN. El objetivo principal de este trabajo se basa en el trabajo de Mena, R. & Walker, S. (2009), en el que se proporciona una idea descollante en la que a partir de una descomposición latente, se puede obtener fácilmente las transiciones así como simulaciones de los tres procesos fundamentales de Ornstein-Uhlenbeck, Cox-Ingersoll-Ross y Wright-Fisher. Es decir, se utiliza una idea modular de descomposición latente para que a partir de tres modelos muy populares: Normal-Normal, Gamma-Poisson y Beta-Binomial se obtengan representaciones equivalentes de los modelos Ornstein-Uhlenbeck, Cox-Ingersoll-Ross y Wright-Fisher, respectivamente; al encontrar formas fun-

cionales de parámetros que dependen del tiempo que hagan que las transiciones satisfagan Chapman-Kolmogorov. Esta descomposición hace que las funciones de transición se conceptualicen de otra forma (sin distracciones numéricas asociadas al cálculo de las funciones especiales) y la simulación de estos procesos de difusión. Esta construcción también hace que las condiciones para la existencia de soluciones vistos a partir de sus ecuaciones diferenciales estocásticas subyacentes sean secundarias (condiciones que eran muy importantes en la estimación de parámetros de EDEs tradicional). Además esta construcción también hace que las condiciones con respecto a los parámetros de la ecuación diferencial estocástica para la existencia de una distribución estacionaria también pasen a segundo plano (condiciones que eran fundamentales con las técnicas de inferencia tradicionales). Esto permite que para un conjunto de puntos en el tiempo (incluso no necesariamente equidistantes) observados en forma de serie de tiempo se pueda trabajar con una verosimilitud aumentada (aumentada por la variable latente de la descomposición latente) e implementar de manera “sencilla” algoritmos de esperanza-maximización para generar sucesiones de parámetros que converjan al parámetro (vectorial) máximo verosímil.

7.3 Temas Selectos de Minería de Textos

Sergio Daniel Raya Rios (CIMAT)

RESUMEN. La minería de textos ha ganado especial importancia en los últimos años dada la gran cantidad de datos que se han generado en la web a través de las redes sociales, las revistas y periódicos digitales, blogs y en general cualquier aplicación que genere texto. El objetivo de este trabajo ha sido investigar algunas de las técnicas más importantes que se proponen dentro de la minería de textos y ejemplificarlas mediante sus aplicaciones a problemas que pueden surgir en diversos contextos. Una aplicación referente al análisis de sentimientos se presenta en este trabajo y tiene como objetivo analizar un conjunto de opiniones que han sido recolectadas de Amazon México sobre diversos productos e identificar patrones útiles en ellas que puedan ser aprovechadas en un contexto de negocio. Dada la flexibilidad de la modelación bajo el enfoque Bayesiano no paramétrico se han considerado los modelos de tópicos latentes (Blei et al. 2003) en esta aplicación. Otro problema clásico en la minería de textos es la detección de SPAM. Este problema ha sido abordado desde muy diversas perspectivas y en este trabajo se ofrece un tratamiento estadístico como parte de las aplicaciones para resolver el problema de la detección de SPAM en teléfonos móviles. Los algoritmos considerados para esta tarea han sido máquinas de soporte vectorial, redes neuronales y clasificadores Bayesianos, entre otros. Es difícil imaginar qué otras cosas pueden ser más difíciles de modelar para un estadístico además de los sentimientos y los pensamientos humanos que pueden ser plasmados en forma de texto. A pesar de que el cerebro ha sido uno de los órganos más estudiados y fascinantes no ha terminado de comprenderse. En el mejor de los casos nos tendremos que defender ante las críticas a nuestros intentos por modelar su comportamiento con la famosa frase “Todos los modelos son erróneos, pero algunos son útiles” que en algún momento mencionó el proclamado estadístico George Box.

7.4 Modelos de Markov ocultos para describir los niveles de ozono en la atmósfera de la Ciudad de México
Araceli Ramírez López (Universidad Autónoma Metropolitana)

RESUMEN. Un modelo de Markov oculto (HMM) es un doble proceso estocástico con un proceso subyacente que no es observable (oculto) pero que puede ser inspeccionado a través de otro proceso estocástico que genera una secuencia de observaciones; en otras palabras, se trata de un proceso bivariado que incluye una cadena de Markov, donde cada uno de sus estados tiene asociada una distribución de probabilidad. En esta intervención se presenta de manera breve y sencilla la estructura matemática de los modelos de Markov ocultos; en particular se describe los tres problemas que plantea este modelo: estimación de parámetros, estimación de la secuencia de estados más probable y evaluación del ajuste del modelo. En particular, se describe la aplicación de los modelos de Markov ocultos a la serie de tiempo de los niveles de ozono en la atmósfera de la Ciudad de México; esta consiste en establecer una relación entre los estados de la cadena de Markov y los regímenes de concentración del contaminante, así como una relación entre la distribución de las concentraciones del contaminante en cada régimen y la distribución asociada a cada estado de la cadena de Markov. Se comentará sobre la estimación de los parámetros del modelo, los resultados y su interpretación. En particular, se proporcionara información sobre el primer régimen de contaminación o contaminación de fondo, un indicador de la exposición real de las personas al contaminante.

CONTRIBUCIONES LIBRES

PONENCIAS

MODELOS LINEALES

Moderador: **Nélida Susana Ozán** (Universidad Nacional de San Juan)

Tests of mean-variance efficiency using the multivariate t-distribution: An application to Latin America

Manuel Jesús Galea Rojas (Pontificia Universidad Católica de Chile)

RESUMEN. In this work we consider the Capital Asset Pricing Model under the multivariate t-distribution with finite second moment. This distribution, which contain the normal distribution, offer a more flexible framework for modeling asset returns. The main goal of this study is to consider the tests of mean-variance efficiency on a given portfolio using the likelihood-ratio, Wald, score and gradient statistics. We provide analytical expressions for the score function and Fisher information matrix. The results are illustrated by using a set of six markets in Latin American countries. The mean-variance efficiency of the EM index is analyzed using an International Capital Asset Pricing Model. Our main conclusion is that the t-distribution shows a better fit and that the mean-variance efficiency of the EM index cannot be rejected.

Modelo de ecuaciones estructurales en estudios de opinión

Indra del Socorro Juárez Vera (Universidad Veracruzana)

RESUMEN. Hoy en día los cuestionarios (instrumentos) se han vuelto de gran importancia dentro del ámbito de los estudios de opinión, ya que se consideran una herramienta útil para la recolección de información de la opinión pública. Los cuestionarios se conforman por ítems que a su vez se encuentran en diferentes escalas de medición, no obstante, comúnmente se utiliza la escala de tipo Likert (ordinal), la cual se comprende de categorías con tendencias positivas, negativas y neutras, sin embargo, estas categorías deben ser evaluadas por diversos métodos o técnicas estadística que permitan ver si son medidas de forma correcta, haciendo necesario el proceso de validación de un instrumento de medición. La validación de cuestionarios es fundamental, puesto que sin ella se desconocería el grado de acuerdo o relación de los ítems de un instrumento, es decir, si el instrumento mide realmente lo que se requiere, así como también la consistencia interna del mismo. Uno de los métodos más utilizados para el proceso de validación es el Alpha de Cronbach, el cual en muchas ocasiones presenta susceptibilidad por el tipo de escala empleada en cada ítem, por lo que los Modelos de Ecuaciones Estructurales muestran una gran alternativa para contrastar la validez de un instrumento. Estos modelos son una técnica multivariante que proporciona suficiente evidencia estadística para identificar la verdadera dimensión de las escalas consideradas en el cuestionario. El presente trabajo de investigación tiene como objetivo principal mostrar la funcionalidad de los Modelos de Ecuaciones Estructurales en los estudios de opinión aplicado a un caso de estudio, describiendo la definición y aplicación de los modelos en el software R-project.

ANÁLISIS DE SUPERVIVENCIA

Moderador: **Alan Riva Palacio Cohen** (University of Kent)

Un modelo de eventos recurrentes aplicados a condiciones de repitencia y deserción universitaria

César Augusto Serna Mejia (Universidad Central)

RESUMEN. Un modelo de eventos recurrentes aplicados a condiciones de repitencia y deserción universitaria. Introducción: Tradicionalmente se ha hablado del problema de la deserción como un mal que aqueja a los diversos niveles de la academia, sin embargo es poco lo que se sabe acerca de sus verdaderos orígenes, los cuales tienen múltiples procedencias. Uno de los causales más frecuentes de deserción universitaria es la repitencia académica; fenómeno académico que ocurre cuando un estudiante no logra, en primera instancia, aprobar satisfactoriamente uno de los cursos en su plan de estudio. Como la repitencia es una condición que puede suceder más de una vez, a la hora de cursar una carrera universitaria, es un escenario ideal para implementar un modelo de recurrencias, técnica estadística que hace parte de la familia de los modelos de supervivencia. Objetivo General: Implementar un modelo de eventos recurrentes para caracterizar la repitencia académica universitaria, modelo útil para calcular la propensión (scoring) a la repitencia, bajo distintos escenarios. Hallazgos: Se utilizó como insumo, información de estudiantes matriculados en diferentes programas de las tres facultades de la Universidad Central, observados en lapsos de tiempos de 11 semestres, en cada una de las seis cohortes entre 2007-I y 2009-II. Se evaluaron características académicas y sociodemográficas del ente implicado, donde se pudo apreciar que para la variable constitución del hogar, el hecho de no vivir con ambos padres o vivir solo, son factores que afectan en gran medida el desarrollo académico de los universitarios, especialmente en los inicios de la carrera. La implementación de este modelo señaló también, estar asociada con la ocupación de la madre, sin dejar a un lado la educación de ambos padres, por tanto, induce a presentar un alto porcentaje de recurrencia; esto entre muchos otros hallazgos.

Estimación de los niveles de fecundidad en México, una aplicación de los modelos de supervivencia y muestreo bootstrap

Leonardo Ailines Genis (IIMAS-UNAM)

RESUMEN. El presente trabajo tiene como objetivo estimar las tasas de fecundidad en México con datos de la Encuesta Nacional de la Dinámica Demográfica (ENADID 2014) a partir de la exposición al riesgo de embarazo y cuyo resultado será un hijo nacido vivo. Para ello se plantea ajustar la exposición al riesgo con un modelo de supervivencia, en donde se define como u^* al tiempo transcurrido a partir de que una mujer tuvo a su último hijo nacido vivo, T como el límite superior del intervalo de

tiempo de estudio $[0, T]$, así como la edad de la mujer al momento del levantamiento de la encuesta. El estimador de la tasa de fecundidad se definirá como el total de mujeres que tuvieron a un hijo nacido vivo dentro del periodo de estudio entre el total de mujeres que fueron expuestas al riesgo de embarazo en el mismo intervalo de tiempo. Dado que se está trabajando con una encuesta de diseño complejo es necesario establecer las precisiones estadísticas de los resultados obtenidos, por ello se complementará la información obtenida con el cálculo de los intervalos de confianza y del coeficiente de variación para cada una de las tasas obtenidas a partir de un muestreo por bootstrap.

La información será desglosada por entidad federativa, lugar de residencia, escolaridad y condición de habla de lengua indígena. Esta información permitirá analizar el desempeño de las políticas públicas en el tema de planificación familiar, la ideología de las familias sobre el número ideal de hijos que desean tener, el acceso a la información sobre el uso de métodos anticonceptivos así como la cobertura en los grupos con una mayor vulnerabilidad.

MARTES 2 DE OCTUBRE
11:00 A 12:00 HRS

PONENCIAS (CONTRIBUCIONES LIBRES)
AUDITORIO ANTONIO ALATORRE

ANÁLISIS MULTIVARIADO

Moderador: **María Teresa Ortiz Mancera (ITAM)**

Caracterización de la violencia en el noviazgo entre adolescentes de dos instituciones universitarias

Jose Giovany Babativa Marquez (Universidad Santo Tomás)

RESUMEN. A diferencia del problema de la violencia conyugal contra las mujeres y de la que se ejerce contra los niños, el problema de la violencia en el noviazgo ha sido poco estudiado, sobre todo en países de Latinoamérica” (Castro y Casique, 2010). Con el objetivo de contribuir al conocimiento sobre el tema, se identifican y analizan las actitudes y conductas agresivas en el marco de las relaciones de noviazgo de estudiantes de dos instituciones universitarias de la ciudad de Bogotá. Las actitudes y conductas se enmarcan dentro de lo que se puede denominar como violencia en el noviazgo entendida como todo acto, omisión, actitud o expresión que genere, o tenga el potencial de generar daño emocional, físico o sexual a la pareja afectiva con la que se comparte una relación íntima sin convivencia ni vínculo marital.

Para el análisis de los datos se usaron técnicas estadísticas multivariadas como el análisis de correspondencias múltiple y el análisis clúster, lo cual permitió clasificar y perfilar las conductas de los adolescentes en su relación así como las variables más influyentes desde su entorno familiar. Adicionalmente, un modelo de regresión multinomial es ajustado con la información de la encuesta, lo que permitió encontrar factores de riesgo asociados con la violencia en el noviazgo.

MARTES 2 DE OCTUBRE
11:00 A 12:00 HRS

PONENCIAS (CONTRIBUCIONES LIBRES)
AUDITORIO MATUTE REMUS

ESTADÍSTICA BAYESIANA

Moderador: **Luis Enrique Nieto Barajas** (Instituto Tecnológico Autónomo de México)

Medidas de probabilidad aleatorias basadas en conjuntos aleatorios

María Fernanda Gil Leyva Villa (IIMAS)

RESUMEN. Se propone usar medidas de probabilidad aleatorias basadas en conjuntos aleatorios para la estimación de densidades, la estimación puede realizarse a un tiempo fijo o bien cuando se tiene una colección de densidades que dependen del tiempo. Además de explorar una construcción novedosa, presentamos un método de estimación y simulación, mismo que probamos con datos simulados

Mezclas de distribuciones con pesos ordenados

Asael Fabian Martínez Martínez (Universidad Autónoma Metropolitana)

RESUMEN. Los modelos de mezclas son una herramienta poderosa y flexible para modelar heterogeneidad, además de ser ampliamente utilizados en diferentes áreas. Bajo un enfoque bayesiano no paramétrico, la distribución mezcla está dada por una medida de probabilidad aleatoria (MPA) discreta, y el modelo resultante posee propiedades interesantes ya que, por ejemplo, permite hacer inferencia sobre el número de grupos, así como sobre la estructura de agrupamiento subyacente en los datos. Entre las diferentes MPAs, el proceso Dirichlet es el más conocido y utilizado. Sin embargo, existe una nueva clase de medidas de probabilidad, que tienen su origen en el denominado proceso geométrico, los cuales tienen una estructura más simple (sin perder flexibilidad) en sus pesos, haciéndolas una opción atractiva. En este trabajo, se presentarán ejemplos específicos de estos modelos de mezclas con pesos decrecientes, estudiaremos sus propiedades e ilustraremos su desempeño en análisis de conglomerados y estimación de densidades. Asimismo, compararemos estos modelos con aquel obtenido del proceso Dirichlet.

ESTADÍSTICA BAYESIANA

Moderador: **Javier Santibáñez Cortés** (Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas)

Conteo rápido en las elecciones

María Teresa Ortiz Mancera (ITAM)

RESUMEN. En este trabajo se presenta una de la metodologías estadísticas que se utilizó para anticipar los resultados finales de las elecciones del 2018. En México, el procedimiento para estimar el porcentaje de votos a favor de los candidatos en el día de la elección se conoce como conteo rápido y es organizado por la autoridad electoral. El conteo rápido consiste en seleccionar una muestra aleatoria de las casillas de votación y analizar sus resultados. El análisis puede hacerse con distintas metodologías, nosotros presentaremos un modelo de postestratificación con regresión multinivel, que es un método bayesiano para corregir diferencias conocidas entre la muestra y las poblaciones objetivo.

Factores asociados al desempeño en matemáticas en estudiantes de bachillerato, Planea 2017.

Elizabeth Solís Alonso (Universidad Autónoma de Guerrero)

RESUMEN. En México, el desempeño en Matemáticas se ha medido y estudiado por diversos organismos nacionales e internacionales como TIMSS, PISA, CENEVAL, ENLACE y recientemente PLANEA. PLANEA Media Superior es una prueba estandarizada cuyo objetivo es medir conocimientos y habilidades de estudiantes mexicanos con el uso de aprendizajes clave. La prueba consta de 100 reactivos de opción múltiple, 50 que evalúan Lenguaje y Comunicación y 50 que evalúan matemáticas, esto genera información diagnóstica del último ciclo de bachillerato. Los resultados de esta prueba dan evidencia del problema de bajo desempeño de los estudiantes de bachillerato en matemáticas en México.

El objetivo de este trabajo es modelar la relación entre el desempeño en matemáticas y variables contextuales en estudiantes de bachillerato en México. Los datos se obtuvieron de la prueba PLANEA 2017 y se utilizan modelos multinivel bajo los enfoques estadísticos clásico y bayesiano.

ESTADÍSTICA EN LA INDUSTRIA

Moderador: **José Edgar Alcántar Muñoz** (Universidad Juárez del Estado de Durango)

Mezcla finita de procesos de degradación inversos Gaussianos

Abelardo Montesinos López (Universidad de Guadalajara)

RESUMEN. Gracias al desarrollo tecnológico, con el paso del tiempo se tienen productos manufac-turados de una confiabilidad cada vez mayor, lo cual hace difícil la evaluación de su confiabilidad usando pruebas de vida tradicionales, pues durante un experimento típico se observa un número reducido de fallas.

Por ello, debido a la necesidad de tomar decisiones en tiempos cortos para éste tipo de productos, una alternativa es el uso de datos de degradación cuando existe alguna característica de calidad de éste cuya degradación a través del tiempo puede relacionarse a su confiabilidad. Aún con pocas fallas y error de medición presente, el análisis de degradación produce estimaciones más precisas debido a que toma en cuenta la cantidad de degradación de las observaciones censuradas modelando directamente ésta con el tiempo.

Los datos de degradación usualmente se analizan con modelos de trayectorias de degradación ge-nerales que por medio de efectos aleatorios toman en consideración la variabilidad de las unidades. Otra aproximación muy común es el uso de procesos estocásticos. Por ejemplo, el proceso inverso Gaussiano (PIG). Sin embargo, en algunas situaciones éste no es suficiente para capturar la variabili-dad entre las unidades ya que sus parámetros pueden variar de unidad a unidad, a causa de diferentes factores. Para tomar en cuenta esto, recientemente se han propuesto extensiones de éste modelo que básicamente son mezclas continuas de un PGI, donde sus parámetros se consideran efectos aleatorios.

En vez de una mezcla continua de procesos inversos Gaussianos, aquí proponemos usar una mezcla finita de éstos para modelar la heterogeneidad de trayectorias de degradación. La estimación de los parámetros la hacemos por máxima verosimilitud empleando el algoritmo EM, se da una ilustración empleando un conjunto de datos, y la contrastamos con un modelo reciente.

Conocimiento a partir de Text Mining y facturación electrónica

Ana Lucía Oña Macías (Universidad de Atacama)

RESUMEN. En los últimos años la generación de información ha crecido sustancialmente. La tec-nología, la globalización y los nuevos avances hacen que muchas instituciones, tanto públicas como privadas, registren diariamente millones de transacciones. Esta información contribuye directamen-te en la toma de decisiones, sin embargo, los usos de estos datos aún están siendo descubiertos. Las Administraciones Tributarias mediante la facturación electrónica logran capturar a nivel de producto

las compras y ventas de personas y empresas tanto nacionales como extranjeras que se llevan a cabo en sus territorios. Esta información que se genera por una necesidad medioambiental, de eficiencia y de control tributario puede llegar a ser empleada para tener datos descentralizados muy valiosos para análisis económicos y sociales. Mediante el uso del Text Mining, esta investigación muestra los posibles usos de los datos generados por la facturación electrónica para generar conocimiento a partir de análisis asociados con preferencias de los consumidores, estadísticas sectoriales, indicadores económicos e información pertinente y focalizada a distintas áreas.

MARTES 2 DE OCTUBRE
16:00 A 17:00 HRS

PONENCIAS (CONTRIBUCIONES LIBRES)
AUDITORIO ANTONIO RODRÍGUEZ

ESTADÍSTICA NO PARAMÉTRICA

Moderador: **Manuel Jesús Galea Rojas** (Pontificia Universidad Católica de Chile)

Proceso Dirichlet compuesto

Arrigo Coen Coria (Facultad de Ciencias)

RESUMEN. Esta plática se enfoca en el comportamiento de sumas aleatorias intercambiables bajo esquemas de intercambiabilidad no-paramétricos donde el número de sumando es dirigido por un proceso Dirichlet. Este tipo de estructuras nos permite poner distintos regímenes de dependencia para así mejorar la semejanza de los modelos con el fenómeno que se está estudiando. Se presentarán los resultados de una fórmula general para el cálculo de dichas sumas, al igual que expresiones explícitas las familias de distribuciones exponencial, gausiana, y tipo fase. También se presentará algoritmo para su cálculo.

*Comparing Bootstrap and asymptotic trimmed runs test
for symmetry in samples coming from the GLD*

Jimmy Antonio Corzo Salamanca (Universidad Nacional de Colombia)

RESUMEN. We propose a bootstrap trimmed runs test for the location alternative, in samples coming from the Generalized Lambda Distribution (GLD) with unknown median, and by mean of Monte Carlo study we compare its empirical powers to identify which of them show better behavior. We also compare their empirical powers with those of the tests of Cabilio & Massaro (1996), Miao et. al (2006), Mira (1999) and with a modification of the Welch(1938) test, all of them for the same purpose. We calibrate the size of the test with samples coming from six symmetrical cases of the GLD and show that the proposed test is empirically unbiased in all calibration cases. Results show that both tests are most powerful than some of the compared tests.

References

- Cabilio, P. & Masaro, J. (1996), "A simple test of symmetry about an unknown median", *The Canadian Journal of Statistics* 24(3), 349-361.
- Miao, W., Gel, Y. & Gastwirth, J. (2006), "A new test of symmetry about an unknown median", *Random Walk, Sequential Analysis and Related Topics - A Festschrift in Honor of Yuan-Shih Chow*. Eds.: Agnes Hsiung, Cun-Hui Zhang, and Zhiliang Ying, World Scientific Publisher pp. 1-19.
- Mira, A. (1999), "Distribution-free test for symmetry based on Bonferroni's measure", *Journal of Applied Statistics* 26(8), 959-972.

Welch, B. L. (1938), "The significance of the difference between two means when the population variances are unequal", Biometrika 29, 350-362

ESTADÍSTICA Y FINANZAS

Moderador: **Jorge Andrés González Burgos** (Pontificia Universidad Católica de Chile)

Volatility analysis through a deep neural stochastic model

Marvin Isaac Querales Carrasquel (Universidad de Valparaíso)

RESUMEN. Volatility is an important indicator of market risk. It is usually expressed as a percentage and calculated as the deviation recorded by an asset with respect to the average of its historical price on a given period. Classical mathematical methods to study volatility involve time series measurements whose behavior has a stochastic characteristic because both the mean and the variance are non-constant. The most outstanding models are the autoregressive conditional heteroskedasticity (ARCH) model and the generalized ARCH (GARCH), which incorporate the functional relationships that allow to relate the current conditional volatility with the past conditional volatilities assuming conditions and previous assumptions. Otherwise, Heston model is a generative model of underlying dynamics and latent variables that provides a closed-form solution for options pricing. Taking advantage of deep learning, Luo et al proposed a neural network of reformulated stochastic volatility in which the GARCH model and the Heston model take place. The objective of this study was to apply this neuronal network in a set of economic data to predict volatility. We simulated 2041 observations of volatility under t-student distribution for 3 degrees of freedom and data was multiplied by a smoothing function of the volatility with a window of 125 observations. Despite data had a high noise, the overall performance of NSVM showed a good prediction and high accuracy of the mean, standard deviation and its confidence bounds. This implies that neural network has a good training and it could be an useful tool to predict volatility.

Estimación robusta de la frontera eficiente para portafolios de inversión

Eduardo Nelson Gutierrez Turner (Universidad del BíoBío)

RESUMEN. Obtener la mayor rentabilidad con el menor riesgo requiere diversificar las inversiones, esto es invertir en un portafolio de acciones compuesto por un conjunto de activos a los que se le asigna un porcentaje de acuerdo a un vector de pesos. El enfoque de Markowitz es la herramienta clásica usada para estimar el vector de pesos y así encontrar la frontera eficiente entre riesgo y rentabilidad. Este enfoque se sustenta en una buena estimación inicial del vector de medias y de la matriz de covarianzas. El método de máxima verosimilitud es ampliamente utilizado, sin embargo es muy sensible ante la presencia de datos atípicos. Los métodos robustos usados hasta el momento (minimum covariance determinant y minimum volume ellipsoid) sufren el problema del enmascaramiento, sin embargo un nuevo método propuesto por Grossi y Laurini, basado en la búsqueda hacia adelante promete obtener mejores resultados. A través del estudio de simulaciones de Monte Carlo

y aplicaciones a datos reales de la bolsa de valores de Chile se muestran las comparaciones entre el método clásico de máxima verosimilitud y los métodos robustos, así también las ventajas de la estimación basada en la búsqueda hacia adelante. Finalmente se presenta en lenguaje de programación R los comandos que se utilizaron para realizar las simulaciones y las aplicaciones.

ANÁLISIS MULTIVARIADO

Moderador: **Miguel Yáñez Alvarado** (Universidad del Bío-Bío)

Alternativas al K-means adaptadas a datos masivos

Ana Belém Nieto Librero (Universidad de Salamanca)

RESUMEN. En la actualidad, los recientes avances en tecnología, los grandes repositorios de datos y el uso de las TICs, han generado bases de datos formadas por miles de variables. Al considerar un número elevado de variables, su análisis no es simple y directo. Además, el gran volumen de datos generados no refleja por sí mismo toda la información que contienen. Para poder extraer conocimiento de ellos son necesarias técnicas de análisis capaces de manejar datos masivos. Estas técnicas deben aprovechar toda la información disponible para generar una buena reproducción de la realidad, pues la comprensión de la mayor parte de los fenómenos radica en la visión multivariante del mundo. Uno de los objetivos de la minería de datos es el uso de técnicas automáticas de reconocimiento de patrones para detectar una estructura de grupos subyacentes a los datos. El objetivo principal de las técnicas de clasificación consiste en asignar un objeto para un grupo subyacente dentro de los datos multidimensionales originales. De las técnicas disponibles, el método de clúster más utilizado es el K-means. Sin embargo, presenta algunos inconvenientes como la aleatoriedad en la inicialización o que el número de grupos debe ser conocido a priori. Todo ello ha hecho que, en los últimos años, diversos autores hayan propuesto nuevas alternativas. Por un lado, aparece el método de agrupamiento K-MEDOIDS, menos sensible a valores atípicos que el K-means y sus extensiones CLARA y CLARANS. Otra de las desventajas del K-MEANS es la forma de los grupos obtenidos. Esta desventaja es mejorada mediante los métodos de agrupamiento basados en densidades, como DBSCAN. Por otro lado, la literatura recoge otros algoritmos novedosos de la familia K-means, como K-means ++, Fuzzy C-Means o Kernel K-Means. En este trabajo se describe la revisión y aplicación de algunos de estos métodos, comparando los resultados obtenidos entre sí.

Análisis de componentes principales geográficamente ponderado como herramienta en el análisis de la sostenibilidad

Nathalia Diazibeth Tejedor Flores (Universidad de Salamanca)

RESUMEN. El Nexo Agua-Energía-Alimentos se basa en una visión integral de la sostenibilidad donde se trata de establecer el equilibrio entre las necesidades de las personas y del medio ambiente. La metodología del Análisis Integrado Multi-Escala del Metabolismo de la Sociedad y el Ecosistema (MuSIASEM, por sus siglas en inglés) permite analizar eficazmente el nexo entre el agua, la energía y los alimentos, por lo cual el objetivo principal de esta investigación es enriquecer esta metodología con información del análisis de datos multivariantes para mejorar la eficiencia de los modelos existentes

de sostenibilidad. Dentro del Análisis Multivariante se aplicó el Análisis de Componentes Principales Geográficamente Ponderadas (GWPCA, por sus siglas en inglés) con la finalidad de tener en cuenta la heterogeneidad espacial de la base de datos. Se utilizó un caso práctico basado en números tomados de situaciones del mundo real y relacionado con el análisis de la sostenibilidad de las sociedades humanas. Para ello se evaluó el Nexo Agua-Energía-Alimentos en Ecuador. Se encontró que existe una variación geográfica en los datos de porcentaje de varianza explicada para las 3 primeras componentes principales; con porcentajes más altos situados en el sur-oeste (región de la costa) y en una pequeña parte de noreste (región de amazónica). Con el GWPCA se observó una clara variación geográfica en la influencia de cada variable en la primera componente. Por ejemplo: la suma de toda la energía que se utiliza en el trabajo remunerado domina en el sur-oeste (región de la costa), mientras que la variable de cantidad de energía utilizada por hora domina en la parte norte de Ecuador. Se concluye el PCA puede reemplazarse por el PCA geográficamente ponderado (GWPCA), cuando queremos tener en cuenta la heterogeneidad espacial y utilizarse como herramienta para conocer la sostenibilidad de un país o región.

ACP en presencia de datos faltantes

Victor Manuel Gonzalez Rojas (Universidad del Valle)

RESUMEN. Se presenta un método para el Análisis de Componentes Principales sobre una matriz de datos disponibles estandarizados Z contenido datos faltantes. Se apela al principio de datos disponibles como en NIPALS y a las relaciones de transición para el cálculo de las matrices $Z'NZ$ y $N^{(1/2)}ZZ'N^{(1/2)}$ cuya descomposición espectral permiten la obtención de los vectores propios $u?R^p y v?R^n$ respectivamente, con los mismos valores propios $?_1, \dots, ?_p$. Con esta propuesta se garantiza la ortonormalidad de los vectores propios y se corrige la falta de ortogonalidad de estos, presente en otras soluciones como las dadas a través de las funciones nipals() de las librerías pls y plsdepot del entorno de programación R.

MARTES 2 DE OCTUBRE
17:30 A 19:00 HRS

PONENCIAS (CONTRIBUCIONES LIBRES)
AUDITORIO ANTONIO RODRÍGUEZ

MODELOS LINEALES

Moderador: **Indra del Socorro Juárez Vera** (Universidad Veracruzana)

*Un enfoque robusto para modelos lineales generalizados via el procedimiento máximo
Lq-verosímil*

Felipe Osorio Salgado (Universidad Técnica Federico Santa María)

RESUMEN. En este trabajo se propone un procedimiento para estimación robusta en el contexto de modelos lineales generalizados (GLM) basado en el método máximo Lq-verosímil. Se presenta un algoritmo de estimación que representa una extensión natural del método de mínimos cuadrados iterativamente ponderados habitual en GLM. Además, obtenemos la distribución asintótica de los estimadores propuestos y se presenta algunos estadísticos de prueba para llevar a cabo ciertas hipótesis de interés. Investigamos la eficiencia de la metodología mediante un estudio de simulación y el análisis de un conjunto de datos previamente analizado en la literatura.

Intervalos de confianza asintóticos para la predicción en regresión con mínimos cuadrados parciales: un estudio de Monte Carlo

Luis Alberto Firinguetti Limone (Departamento de Estadística, Facultad de Ciencias, Universidad del Bío-Bío)

RESUMEN. En el contexto del modelo clásico de regresión lineal este trabajo estudia la construcción de intervalos de confianza asintóticos para la predicción en base al estimador de mínimos cuadrados parciales. Se estudian en las propuestas de Denham [2000], Phatak, Reilly y Pendilis [2002] y Romera [2010]. Debido al carácter no lineal del estimador de mínimos cuadrados parciales, resulta especialmente difícil, si no imposible, el estudio de las propiedades exactas del estimador. Luego para evaluar y comparar el comportamiento de estos intervalos de confianza en muestras pequeñas se realiza un estudio de simulaciones. Para lo cual se implementan en Matlab, los algoritmos necesarios para construir los intervalos de confianza. Una conclusión importante del estudio es que los intervalos asintóticos propuestos por Denham [2000], Phatak, Reilly y Pendilis [2002] son muy razonables, aun en muestras pequeñas, pues presentan probabilidades de coberturas cercanas a las nominales. En cambio la propuesta de Romera [2010] no produce buenos resultados.

Bibliografía

- Denham, M.C. (2000) Choosing the number of factor in partial least squares regression: estimating and minimizing the mean squared error of prediction Journal of Chemometrics, 14 (pp.351-361).
Phatak, A., Reilly P.M., Pendilis A: (2002) The asymptotic variance of the univariate PLS estimator Linear Algebra and its Applications, 354 (pp.245-253).

Romera, R.(2010) Prediction interval in Partial Least Squares regression via a new local linearization approach Chemometrics and Intelligent Laboratory System, 103 (pp.122-128).

MUESTREO

Moderador: **Jimmy Antonio Corzo Salamanca** (Universidad Nacional de Colombia)

Muestras aleatorias con unidades muestrales impuestas: Estado de avance

Ernesto A. Rosa Bonilla (Universidad Nacional de Tres de Febrero)

RESUMEN. Al realizarse un Diseño Muestral, existen situaciones en las que, por fuera de los aspectos técnicos, se solicita a quien realiza el diseño la inclusión en la Muestra de unidades “especiales”, ya sea antes o después de las que deben ser seleccionadas aleatoriamente. Este tipo de situaciones provoca usualmente reacciones adversas del profesional que diseña la Muestra, ya que en apariencia, el acceder a ese requerimiento o exigencia, vulneraría la aleatoriedad que debería imperar en todo Diseño Muestral que se precie. El esbozo de este problema, planteado con un caso particular de Diseño Muestral aplicado a un estudio relativo a servicios asistenciales en salud, fue presentado en el Congreso Interamericano de Estadística realizado en 2017 en R. Argentina, expresándose en forma de “conjetura”, la posible solución que se tendría. Posteriormente se presentó el tema como un Proyecto de Investigación interno de la Universidad Nacional de Tres de Febrero (UNTREF), correspondiendo este Resumen al “estado de avance” del mismo, que incluye la posible solución teórica del problema. Esta solución se basa en dos aspectos del Diseño aplicado:

- Considerar a las unidades muestrales “impuestas” como casos de inclusión forzosa, que en los estimadores y sus dispersiones se representan solo a sí mismas.
- Determinar los factores de expansión de cada una de las unidades seleccionadas al azar, para que representen a la población final, una vez excluidas las agregadas forzosamente.
- Las expresiones de los estimadores, sus dispersiones y los factores de expansión y calibración resultantes, corresponden naturalmente a un Diseño Muestral complejo a varias etapas.

Estimación del parámetro de suavizamiento en muestreo de poblaciones finitas

Luis Fernando Contreras Cruz (Universidad Autónoma Chapingo)

RESUMEN. En el contexto de regresión no paramétrica, en particular, los modelos de splines penalizados incluyen un parámetro de suavizamiento. Existen algunas maneras de obtener dicho parámetro, por ejemplo, validación cruzada. En este trabajo se presenta una forma de estimar el parámetro de suavizamiento bajo el contexto bayesiano. Además, bajo el paradigma de muestreo basado en modelo, se estudia la influencia del parámetro estimado sobre la estimación del total poblacional. Se presenta un estudio de simulación y una aplicación a datos reales.

Optimización multicriterio de la distribución del tamaño de muestra en muestreo estratificado

José Elías Rodríguez Muñoz (Universidad de Guanajuato)

RESUMEN. En el muestreo estratificado se pueden tener varios criterios para asignar el tamaño de muestra en cada estrato. La aplicación de estos criterios puede llevar a varias distribuciones de dichos tamaños de muestra. En este escenario, el usuario de estos criterios se enfrentará a la necesidad de establecer una sola asignación de tamaño de muestra. Dicho usuario puede hacer la asignación única utilizando más información de la población o de forma empírica. En este trabajo se propone un método objetivo para asignar el tamaño de muestra en cada estrato sin utilizar más información de la población. Este método está basado en minimizar el índice de Gini de las posibles distribuciones de muestra comparadas con las distribuciones sugeridas por los criterios de asignación mencionados anteriormente. El resultado de esta asignación óptima se puede interpretar geométricamente a través de la curva de Lorenz de cocientes de tamaños de muestra entre los diferentes criterios.

TEORÍA Y APLICACIÓN DE VALORES EXTREMOS

Moderador: **Arrigo Coen Coria** (Facultad de Ciencias)

Extremos de ozono en nueve regiones la zona metropolitana del valle de México

Sara Rodríguez Rodríguez (Instituto Tecnológico del Altiplano de Tlaxcala)

RESUMEN. Se recopilan observaciones de ozono a partir del año 2000 al 2017 del Sistema de Monitoreo Atmosférico de la Ciudad de México (SIMAT), posteriormente se obtienen los máximos mensuales y se supone que siguen una distribución de Valores Extremos Generalizada, luego se hace un modelo donde se incorpora al parámetro de localización una regresión lineal, con el propósito de obtener un parámetro que indique la tendencia de los máximos de ozono, esta metodología se aplica primero a cada una de las estaciones y luego a nueve regiones geográficas de la Ciudad y Estado de México

Un modelo jerárquico bayesiano para extremos espaciales

José del Carmen Jiménez Hernández (Universidad Tecnológica de la Mixteca)

RESUMEN. La contaminación es un problema que afecta a muchas ciudades del mundo, en muchos casos la teoría espacial de valores extremos se ha usado con buenos resultados. En este artículo proponemos un modelo jerárquico bayesiano espacial para analizar datos máximos de contaminación por monóxido de carbono en la zona metropolitana de la Ciudad de México. Se proponen tendencias espaciales en los parámetros de la distribución de valores extremos y producimos mapas predictivos para los parámetros de la distribución, en estos se puede observar el comportamiento espacial de los parámetros en forma marginal lo cual permite captar la variación local, lo cual no es posible solamente con las tendencias determinísticas.

Teoría de valores extremos en el estudio del nivel de ríos en el estado de Tabasco

Leonardo Alfonso Martinez Gonzalez (Universidad Juarez Autonoma de Tabasco)

RESUMEN. La teoría de valores extremos se encarga del estudio de la probabilidad de ocurrencia de eventos que son más extremos que cualquiera que ya haya sido observado. Esta teoría tiene aplicaciones en diversos campos, entre ellos climatología, medio ambiente, finanzas e ingeniería. En este ponencia se presenta la herramienta clásica de la teoría de valores extremos en el contexto de máximos por bloques de variables aleatorias. También se muestra su aplicación en el análisis de los máximos semanales del nivel de los ríos Grijalva y Usumacinta en el Estado de Tabasco, considerando que los niveles de los ríos son monitoreados cada 10 minutos en sus estaciones automáticas correspondientes.

ANÁLISIS MULTIVARIADO

Moderador: **Alma Janett Tenorio Aguirre** (Universidad Veracruzana)

Una clase de modelos con enlace asimétrico para datos de respuesta binaria

Omar Chocotea Poca (Universidad de Valparaíso)

RESUMEN. Los enlaces simétricos probit y logit son ampliamente utilizados en el modelado de datos de respuesta binaria. Sin embargo, en un escenario de probabilidad extrema el ajuste global puede mejorar significativamente utilizando el enlace asimétrico, por ejemplo, el cloglog. En este trabajo se propone una clase de modelos con enlace asimétrico mejor equipada para el análisis de datos de respuesta binaria utilizando la perspectiva bayesiana. Para la construcción de la clase se dispone de la representación estocástica de Henze (1986) [Henze, N. (1986). A probabilistic representation of the "skew-normal" distribution. Scandinavian Journal of Statistics, 13(4), 271–275.] y la distribución mezcla en escala de la normal asimétrica de Branco & Dey (2001) [Branco, M. D. & Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. Journal of Multivariate Analysis, 79(1), 99–113.]. La clase es atractiva en los siguientes aspectos: a) por permitir regular la asimetría y/o pesadez de las colas de la función de enlace; b) por basarse en el punto de vista de variable latente de Albert & Chib (1993) [Albert, J. H. & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association, 88(422), 669–679.], donde la variable latente tiene una estructura de modelo de efectos-aleatorios con distribución asimétrica; y c) por englobar exactamente o aproximadamente a una variedad de modelos de enlace asimétrico y simétrico. Nuestros estudios con datos simulados y el estudio con datos reales sugieren que, el análisis se puede aplicar a cualquier conjunto de datos de diversa complejidad.

Análisis no paramétrico de curvas ROC

Carlos Cuevas Covarrubias (Universidad Anáhuac, Centro de Investigación en Estadística y Matemáticas Aplicadas)

RESUMEN. La curva ROC constituye una herramienta fundamental para la evaluación de criterios de clasificación estadística definidos en un contexto supervisado. Dada una regla de clasificación, la curva ROC resume describe su desempeño al expresar la sensibilidad como función de la tasa de falsos positivos. Esta información gráfica suele resumirse a partir del área bajo la curva ROC. Proponemos un nuevo método de estimación basado en kérneles Gaussianos de suavizamiento. Este nuevo enfoque permite estimar el área bajo la curva ROC con menor error cuadrático medio. Los conceptos desarrollados se ilustran con ejemplos prácticos basados en datos reales.

ESTADÍSTICA Y SALUD PÚBLICA

Moderador: **Carlos Díaz Avalos** (Universidad Nacional Autónoma de México)

Modelos para mezclas de distribuciones y su aplicación en salud pública

Belem Trejo Valdivia (Instituto Nacional de Salud Pública)

RESUMEN. Los modelos lineales generalizados (glm) permiten describir el comportamiento de la media de una variable respuesta a través de un predictor lineal y una liga apropiada. Este abordaje es útil cuando la distribución de la variable respuesta es unimodal y la media representa adecuadamente la localización. En área de salud (y en otras áreas) existen variables cuya distribución en diferentes poblaciones resulta bimodal o incluso multimodal; esta característica se presenta tanto para variables continuas como discretas. También existen variables (por lo general continuas) cuya distribución es unimodal, pero que ni ellas en su escala original ni alguna transformación de las mismas, puede ser identificada como parte una familia distribucional simple. En esta plática se discute el abordaje vía modelos para mezclas de distribuciones (fmm por sus siglas en inglés) para analizar el tipo de situaciones arriba mencionadas. Se ejemplifica dicho abordaje con el análisis de dos proyectos, uno para el caso de una variable continua (ingesta calórica diaria de niños) y el otro para una discreta (puntaje de una escala de violencia en mujeres).

Uso del SEM en estudios de evaluación educativa a gran escala: efecto de lo económico en el logro dada la intervención de la convivencia escolar.

Violeta De la Huerta Contreras (Instituto Nacional de Evaluación para la Educación)

RESUMEN. Los modelos de ecuaciones estructurales son ampliamente utilizados para explicar fenómenos en las ciencias sociales, debido a su versatilidad de incluir variables observadas y latentes. En estudios de evaluación educativa a gran escala, el objetivo es reportar resultados a nivel poblacional, a partir de datos que provienen de un diseño muestral complejo y la variable de logro educativo construida mediante la metodología de valores plausibles.

El estudio contribuye al desarrollo metodológico de la aplicación del SEM (Structural Equation Modeling). Para ello, se requiere también de otro aspecto fundamental, la teoría del fenómeno educativo. El cual define las variables del modelo y sus relaciones como son el efecto de las condiciones económicas sobre el logro educativo con la convivencia escolar como variable interviniente. Los resultados se reportan a nivel nacional y por los tipos de escuelas secundarias en México con fines comparativos.

Los datos corresponden a la evaluación de matemáticas de tercero de secundaria del Plan Nacional para la Evaluación de los Aprendizajes (PLANEA) 2015 realizado por el Instituto Nacional para la Evaluación de la Educación (INEE). La estimación de los modelos fue realizada en Mplus 8.

ESTADÍSTICA BAYESIANA

Moderador: **Luis Alberto Gutiérrez Inostroza** (Pontificia Universidad Católica de Chile)

Colombian women's life choices: a Bayesian nonparametric multivariate regression approach

Isadora Antoniano Villalobos (Bocconi University)

RESUMEN. Women in the Latin America and Caribbean countries, face difficulties related to the patriarchal traits of their society. In Colombia, the well-known conflict afflicting the country since 1948, has increased the risk of vulnerable groups. It is important to determine if recent efforts to improve the welfare of women have had a positive effect, one that extends beyond the capital, Bogota. In an initial effort to shed light on this matter, we analyze cross-sectional data arising from the Demographic and Health Survey Program which collects and disseminates data on random samples of households selected from a national sampling frame. Our aim is to study the relationship between baseline socio-demographic factors and variables associated to fertility, partnership patterns and work activity. We propose a flexible Bayesian nonparametric multivariate regression model, which can capture nonlinear regression functions and the presence of non-normal errors, such as heavy tails or multi-modality. The infinite mixture model has interpretable covariate-dependent weights constructed through normalization, allowing for combinations of both categorical and continuous covariates, as well as censoring in one or more of the responses. The infinite number of components and the intractable normalizing constant pose computational difficulties for inference. These are overcome through an adaptive truncation algorithm which combines adaptive Metropolis-Hastings and sequential Monte Carlo to create a sequence of automatically truncated posterior mixtures.

Diseño de experimentos bayesiano para pruebas de Tolerancia Oral a la Glucosa (OGTT)

Nicolás Elio Kuschinski Kathmann (Centro de Investigación en Matemáticas)

RESUMEN. Las pruebas OGTT son comúnmente usadas para diagnosticar resistencia a la insulina o diabetes. Nuevos avances en el estudio de pruebas OGTT han hecho que su análisis se plantea como un problema inverso, el cual ha sido estudiado mediante inferencia bayesiana. En base a estas nuevas herramientas de análisis, surge la pregunta de si es posible mejorar la prueba misma. Revisaremos la teoría del diseño de experimentos bayesiano y propondremos un estimador para la utilidad esperada de un diseño. Se estudian las propiedades de este estimador y se usa para generar un m?todo para cuantificar la incertidumbre en comparaciones entre diseños. Este m?todo se usa para encontrar un diseño nuevo, y este diseño se compara favorablemente con el esquema usual.

ESTADÍSTICA EN LA INDUSTRIA

Moderador: **Abelardo Montesinos López** (Universidad de Guadalajara)

Time series analysis in a ceramic tile manufacturing process: a case study

Iván Pacheco Soto (Instituto Tecnológico y de Estudios Superiores de Monterrey)

RESUMEN. The purpose of this study is to improve the statistical process control implemented in a ceramic manufacturing plant in Zacatecas, Mexico to reduce unwanted ceramic tile defects and wastages. Although the ceramic manufacturing process is very complex, many tile defects depend significantly on the raw material, mainly, the clay mixture, potash feldspar and quartz. Furthermore, the heavy use of technology makes these processes almost completely automatized, so it is possible to record many parameters in time. There is empirical evidence and some laboratory experiments that relate many of the tile defects to the powder preparation (powder composition, wet grinding and atomization) and the green body preparation (shaping/pressing and drying), but there is no cross-correlation evidence of these effects. This work presents an analysis and cross-correlation of the series at the different stages of the process, including the firing of the green body. These results can be used to improve the detection of defects in the ceramic tile in the early stages of the process.

Desarrollo de modelo ARX para simulación de la degradación de turbinas de jet bajo condiciones de operación variables

José Edgar Alcántar Muñoz (Universidad Juárez del Estado de Durango)

RESUMEN. Mantenimiento basado en condición (MBC) es una herramienta para calendarizar de manera óptima las tareas de mantenimiento dentro de la industria, impactando directamente en la reducción de costos por el tiempo dedicado a esta actividad y piezas de repuesto, principalmente. MBC presenta la ventaja de monitorear la condición de degradación actual del equipo, con ello busca anticiparse a fallas abruptas. La literatura menciona la necesidad de considerar las condiciones de operación del equipo (carga de trabajo, temperatura, etc.), las cuales son variables durante la vida de éste; sin embargo, pocos estudios han desarrollado metodologías que incorporen dichas condiciones debido al costo que implica la obtención de datos al hacer funcionar un equipo hasta su falla, quienes lo hacen, presentan un limitado número de equipos para su análisis. Este trabajo se centra en desarrollar un modelo estadístico que permita simular tantas degradaciones como se desee, bajo condiciones de operación variables. Se parte del análisis estadístico de una base de datos sobre desgaste en turbinas, perteneciente al repositorio de la NASA, que cumple con las características mencionadas. La investigación realiza un análisis estadístico multivariado a la variable de degradación y condiciones de operación, para luego ajustar un modelo estadístico autorregresivo con variables exógenas (ARX) usando la metodología Box-Jenkins, y establece una distribución de probabilidad multivariada a los

parámetros del modelo. Se obtiene un modelo de simulación donde los parámetros y las condiciones de operación son variables, lo que permite estudiar un número considerable de escenarios de degradación.

ESTADÍSTICA Y SALUD PÚBLICA

Moderador: **Belem Trejo Valdivia** (Instituto Nacional de Salud Pública)

Prevalencia de la obesidad en México: Modelo para variables de entrada

Carlos Díaz Avalos (UNAM)

RESUMEN. En los últimos años, la obesidad se ha vuelto un severo problema de salud en México. Esta enfermedad es asociada como factor de riesgo de hipertensión y diabetes por lo que es necesario analizar la distribución espacial de la prevalencia en orden de comprender el impacto en México. En este trabajo presentamos resultados preliminares usando datos de la Encuesta Nacional de Salud 2012. Nuestra propuesta es ajustar un modelo Poisson jerárquico vía cadenas de Markov para evaluar la prevalencia relativa a nivel de municipios. Los resultados muestran que la prevalencia no tiene una tendencia notable y que existen zonas en México cuya prevalencia es muy alta.

Estudio multivariante de los estilos y estrategias de aprendizaje en los estudiantes de medicina y odontología españoles

María Concepción Vega Hernández (Universidad de Salamanca)

RESUMEN. Los alumnos de los Grados en Medicina y Odontología son un colectivo con una formación muy exigente debido a la cantidad de turnos de prácticas y el desgaste físico; por esta razón, presentan gran interés a la hora de analizar sus estilos y estrategias de aprendizaje. El objetivo de esta investigación se centra en identificar los estilos y estrategias de aprendizaje predominantes en los estudiantes de Medicina y Odontología, analizando sus similitudes y discrepancias. El estudio se llevó a cabo sobre una muestra compuesta por 296 estudiantes de la Universidad de Salamanca (España), pertenecientes al área de Ciencias de la Salud. Para realizar la evaluación de los estilos y estrategias de aprendizaje utilizados por los alumnos se empleó el Cuestionario Honey-Alonso de Estilos de Aprendizaje (CHAEA) de Alonso, Gallego y Honey (1995) y la Escala de Estrategias de Aprendizaje ACRA-Abreviada (ACRA-A) de De la Fuente y Justicia (2003). Se puso de manifiesto que el estilo de aprendizaje predominante en estos estudiantes fue el reflexivo, caracterizados por ser receptivos, capaces de reaccionar, concienzudos, analíticos, cuidadosos y prudentes. No obstante, en el caso de las estrategias de aprendizaje los futuros médicos usan más los hábitos de estudio y los odontólogos las estrategias de apoyo.

ESTADÍSTICA BAYESIANA

Moderador: **Abel Palafox González** (Universidad de Guadalajara)

*Solución numérica del problema inverso de dispersión de ondas en estadística bayesiana
usando representaciones de alto nivel*

Maria Luisa Daza Torres (Centro de Investigación en Matemáticas)

RESUMEN. El problema inverso de dispersión de ondas que abordamos es reconstruir un obstáculo penetrable a partir de observaciones del campo dispersado. Dado que las observaciones están contaminados con ruido, el problema inverso puede considerarse como un problema de inferencia estadística. Formulamos el problema inverso en el marco bayesiano que nos permite introducir conocimiento a priori sobre la naturaleza del obstáculo y a su vez cuantificar la incertidumbre en su reconstrucción. La distribución posterior obtenida se aproxima mediante el algoritmo de Metrópolis-Hastings con un núcleo de transición de probabilidad que conmuta con las transformaciones afines del espacio. Cada muestra de la distribución posterior requiere la solución del problema directo, que es equivalente a resolver la ecuación integral de Lippmann-Schwinger sobre el soporte del obstáculo. En consecuencia, proponemos un método numérico eficiente para resolver la ecuación de Lippmann-Schwinger. El soporte del obstáculo es parametrizado usando una nube de puntos, que es una representación de alto nivel. Las representaciones de alto nivel son intrínsecamente de baja dimensión, es decir, diseñamos un espacio de parámetros de baja dimensión para hacer la inferencia más fácil.

*Control del error en la distribución posterior numérica en el problema inverso de
conductividad térmica en 1D*

José Cricelio Montesinos López (Centro de Investigación en Matemáticas)

RESUMEN. El problema inverso en el campo de transferencia de calor que abordamos es estimar la conductividad térmica a partir de mediciones de la temperatura dentro del cuerpo conductor. Definimos el problema inverso con un enfoque bayesiano y proponemos un campo aleatorio de Markov Gaussiano como distribución a priori de la conductividad térmica. El problema directo lo resolvemos usando el método de elementos finitos, el cual nos permite calcular una cota para el error de la aproximación. A partir de esta cota y del resultado obtenido por C&A, encontramos un tamaño de malla h , para resolver el problema directo, para el cual la posterior numérica sea prácticamente indistinguible a la posterior teórica.

JUEVES 4 DE OCTUBRE
11:00 A 12:00 HRS

PONENCIAS (CONTRIBUCIONES LIBRES)
AUDITORIO ANTONIO RODRÍGUEZ

PROBABILIDAD APLICADA Y PROCESOS ESTOCÁSTICOS

Moderador: **Lizbeth Naranjo Albarrán** (Universidad Nacional Autónoma de México)

Estimación Máximo Verosímil del proceso de riesgo Markov-modulado perturbado por difusión

Fernando Baltazar Larios (UNAM)

RESUMEN. Se presenta un modelo para hacer estimación máximo verosímil para los parámetros del proceso de riesgo perturbado por una difusión y en el que las intensidades con las que se presentan los reclamos, el monto de los reclamos y la volatilidad de la difusión varían en el tiempo de acuerdo a un proceso Markoviano. La propuesta está basada en hacer inferencia de los parámetros involucrados dadas las observaciones de los tiempos y montos de reclamación, así como algunos momentos a tiempo discreto del proceso de riesgo. Dado que se tiene un problema de información faltante, se propone utilizar el algoritmo EM y métodos MCMC para hacer la inferencia.

*Construcción de procesos estacionarios de Markov utilizando las distribuciones
Matriz-Exponenciales Bilaterales*

Luz Judith Rodriguez Esparza (Universidad Autónoma Chapingo)

RESUMEN. En este trabajo se considera una construcción de procesos de Markov con distribución invariantes las distribuciones Matriz-Exponenciales Bilaterales (BME por sus siglas en inglés). Esta clase de distribuciones tienen soporte la recta real y se caracterizan por tener funciones generadoras de momentos racionales, es decir, la razón entre dos polinomios. Estas características de las BME contribuyen a su gran importancia en el área de la modelación estocástica. La estructura de la construcción de estos procesos se basa primeramente considerando las distribuciones de Erlang y Gamma, dando lugar a los procesos BME-ARCH para el caso univariado y los BME-ARCH(q) para el caso multivariado. Luego, se consideran las distribuciones Matriz-Exponenciales, obteniéndose los modelos BME-Normales. Incluimos un estudio de simulación.

DISEÑO DE EXPERIMENTOS

Moderador: **José Cricelio Montesinos López** (Centro de Investigación en Matemáticas)

Análisis del desempeño de un panel para la evaluación de atributos de apariencia y olor de yerbas comerciales utilizando metodología multivariada

Myriam Carmen Nuñez (Universidad de Buenos Aires)

RESUMEN. En los últimos años el sector yerbatero ofrece productos diferenciados, con mayor variedad de sabores incluyendo un porcentaje de numerosas hierbas dando lugar a la categoría de Yerba mate compuesta, y esto produce un sabor diferente al mate tradicional, muy aceptado por los consumidores. Las características del olor y la apariencia de yerba mate son atributos claves en la aceptación del consumidor. El objetivo de este trabajo consistió en analizar la reproducibilidad de un panel, que realizó mediante perfiles libres, la descripción de las similitudes y diferencias sensoriales de un conjunto de yerbas comerciales, elaboradas y compuestas. Para ello se utilizaron 18 marcas comerciales de yerba mate y yerba mate compuesta que fueron evaluadas por un panel de 20 consumidores no entrenados en tres sesiones experimentales. El análisis de la varianza de 3 factores indicó la existencia de diferencias significativas ($p << 0,05$) entre yerbas, para tres atributos analizados: cantidad de palo, tamaño de palo, menta-mentol. No hubo diferencias entre las sesiones, lo cual muestra la consistencia de los jueces. El análisis de las diferencias en las calificaciones de los candidatos estudiados, para cada atributo en las distintas sesiones, indica la concordancia y la capacidad discriminatoria de los jueces para realizar el análisis previsto. El análisis en componentes principales (ACP) demostró que las dos primeras componentes explican el 89,3 % de la varianza total, para el atributo cantidad de Palo (a modo de ejemplo). 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.

Bayesian factor identification in definitive screening designs

Víctor Aguirre Torres (Instituto Tecnológico Autónomo de México)

RESUMEN. Definitive Screening Designs (DSD) are a class of experimental designs that have the possibility to estimate linear, quadratic and interaction effects with relatively little experimental effort. The linear or main effects are completely independent of two factor interactions and quadratic effects. The two factor interactions are not completely confounded with other two factor interactions, and quadratic effects are estimable. The number of experimental runs is twice the number of factors of interest plus one. Several approaches have been proposed to analyze the results of these experimental plans, some of these approaches take into account the structure of the design, others

do not. In this paper we present a Bayesian procedure that incorporates the Bayesian factor identification which is an approach that computes, for each factor, the posterior probability that it is active, this includes the possibility that it is present in linear, quadratic or two factor interactions. This a more comprehensive approach than just testing the significance of an effect.

SERIES DE TIEMPO

Moderador: **Leonardo Moreno Romero** (Departamento de Métodos Cuantitativos, Udelar)

Calibración de modelos matemáticos de crecimiento poblacional utilizando series de tiempo

Francisco Eduardo Novoa Muñoz (Universidad del Bío-Bío)

RESUMEN. En este trabajo se considera el problema de identificación de los parámetros en modelos matemáticos para el crecimiento de una población, los cuales son descritos en términos de ecuaciones diferenciales ordinarias. La identificación se hace a partir de un set de datos reales (experimentales) con ruido. Se parte el estudio del problema de calibración introduciendo una formulación rigurosa como un problema de optimización con restricciones. La función objetivo es la comparación, en el sentido de la norma L₂, entre la solución teórica del modelo matemático y la observación. La restricción es el problema de valores iniciales para la ecuación diferencial que modela el fenómeno en estudio. A continuación, se discute la aproximación numérica del problema de optimización. Primera-miente, se extiende la observación discreta a todo el intervalo de tiempo aplicando series temporales. Luego, se aproxima el modelo matemático por un método numérico y se deduce el gradiente numérico para la función objetivo. Para ilustrar la aplicación del método de calibración propuesto se considera un ejemplo con datos simulados (llamados sintéticos) y un ejemplo con datos obtenidos en el laboratorio.

Modelos GAS aplicados a series temporales climatológicas de Chile

Rodrigo Andrés Gutiérrez Martínez (Universidad del Bío Bío)

RESUMEN. Los modelos GAS (generalized autoregressive score), introducidos por Creal et al. (2008), son modelos con parámetros variables en el tiempo usados para el análisis de series temporales, con el fin de explicarlas y/o hacer predicciones. La actualización de los parámetros variables en el tiempo es dada por el vector score ponderado de la función de verosimilitud. Según la clasificación de Cox (1981), estos modelos pueden ser vistos como modelos guiados por observaciones, en los cuales los parámetros variables en el tiempo son funciones de valores desfazados de la variable dependiente, así como de valores actuales y desfazados de variables exógenas. Para los modelos de la clase GAS, la evaluación de la verosimilitud es bastante sencilla. La incorporación de efectos como asimetría, memoria larga y otras dinámicas también se pueden hacer sin mayores complejidades. Algunos modelos de econometría y finanzas se pueden ver como casos particulares de modelos GAS. Como ejemplos los modelos autorregresivos con heterocedasticidad condicional (GARCH) de Engle y Bollerslev (1986), autorregresivos de intensidad y duración condicional (ACI y ACD), de Engle y Russell (1998) y Russell (2001).

El enfoque GAS puede ser usado para permitir que los parámetros constantes del modelo de espacio estado puedan variar en el tiempo. En particular, se considera la utilización del modelo estructural básico que considera las componentes de nivel, pendiente y estacionalidad con varianzas no constantes.

En el contexto del cambio climático, es importante conocer el comportamiento a lo largo del tiempo del nivel de las series climáticas, en particular las series mensuales de precipitaciones de Chile entre los años 1940 y 2018, enfocados principalmente en la estimación de la componente de tendencia a lo largo del tiempo.

Además, se compara los resultados obtenidos usando el enfoque GAS con los obtenidos del modelo estructural básico de varianza constante.

ANÁLISIS MULTIVARIADO

Moderador: **María Concepción Vega Hernández** (Universidad de Salamanca)

Perfil sociodemográfico de estudiantes universitarios en estudio sobre consumo de drogas legales

Alma Janett Tenorio Aguirre (Universidad Veracruzana)

RESUMEN. Se aplicó por segunda ocasión un estudio sobre el consumo de drogas en jóvenes universitarios en el estado de Veracruz, donde se consideran información sobre prevalencia del consumo de drogas legales e ilegales, factores psicosociales, uso del tiempo libre, autoconcepto, estado de ánimo, relación con los padres y/o pareja, conductas de riesgo, discapacidad entre otras. La base de datos resultante está conformada por más de 200 variables y 19,815 casos, el instrumento aplicado trató de enfocarse en aspectos sociodemográficos de los estudiantes y distinguir aquellas que podrían relacionarse más con el consumo de drogas. La gran cantidad de datos que se obtuvieron en este estudio plantea la necesidad de utilizar técnicas de análisis de datos multivariante, que permita obtener un mejor entendimiento del fenómeno que se presenta sobre el consumo de drogas legales en los estudiantes universitarios. Para poder definir los perfiles sociodemográficos de tal cantidad de información, se aplicó TAID_LCA (Detección automática de interacciones de Tau - Análisis de clases latentes) (Castro y Galindo 2012), análisis de una segmentación estadística con la posibilidad de respuestas multivariadas, bajo el modelo de clase latente. El numerador del índice puede descomponerse a través de los ejes principales, usando el Análisis de Correspondencias No Simétricas (ACNS), (Lauro y Dámbra 1984). Esto permite la segmentación en árboles ternarios a través de la representación en el plano factorial de las coordenadas de las categorías de las variables predictoras. La aplicación de TAID-LCA, permite conocer los perfiles de los estudiantes usuarios y no usuarios del consumo de drogas legales, que además de conocer quienes son estos estudiantes, permitirá elaborar propuestas contextualizadas para tratar de elaborar campañas dirigidas que contemple acciones para la prevención del consumo ocasional, habitual o adictivo de drogas legales.

Ley Tracy-Widom en pruebas de hipótesis para la covarianza poblacional

Didier Cortez Elizalde (Universidad Juárez Autónoma de Tabasco)

RESUMEN. Es bien conocido en análisis multivariado que la estimación de la matriz de covarianza es complicada en el caso en que la dimensión de los datos es mayor o igual al tamaño de la muestra (caso de dimensión alta). En este trabajo se muestra cómo la distribución Tracy-Widom, que aparece en la teoría de matrices aleatorias, puede ser utilizada para llevar a cabo pruebas de hipótesis para determinar si la matriz de covarianza poblacional de datos normales multivariados de dimensión alta es un múltiplo de la matriz identidad o una matriz especificada. También se muestra, mediante

simulaciones, que estas pruebas estadísticas tienen un buen comportamiento, tanto en el caso de dimensión alta como en el caso en que el tamaño de la muestra es mayor a la dimensión de los datos (caso clásico), ya que las proporciones de rechazo bajo la hipótesis nula son cercanas al nivel de significancia considerado.

JUEVES 4 DE OCTUBRE
16:00 A 17:00 HRS

PONENCIAS (CONTRIBUCIONES LIBRES)
AUDITORIO ANTONIO RODRÍGUEZ

ESTADÍSTICA BAYESIANA

Moderador: **Maria Luisa Daza Torres** (Centro de Investigacion en Matemáticas)

Modelización para meta-análisis de test diagnósticos de respuestas binarias: un enfoque desde cópulas y modelos jerárquicos

Johny Javier Pambabay Calero (Escuela Superior Politécnica del Litoral)

RESUMEN. Los procedimientos estadísticos actuales implementados en paquetes de software estadístico para la agrupación de datos de exactitud de pruebas diagnósticas incluyen la regresión de HSROC y el modelo de meta-análisis de efectos aleatorios bivariante (BRMA) . Sin embargo, estos modelos no informan la media global, sino más bien la media de un estudio central y tienen diferencias estimando la correlación entre sensibilidad y especificidad cuando el número de estudios en el meta-análisis es pequeño y / o cuando la varianza entre estudios es relativamente grande.

En el presente trabajo se analizarán los principales métodos estadísticos para meta-análisis de estudios de la exactitud diagnóstica con un mayor énfasis en la modelización conjunta de la sensibilidad y especificidad utilizando cópulas basada en una distribución bivariante beta.

Meta-análisis en DTA con Modelos Jerárquicos Bivariante y HSROC: Estudio de Simulación

Sergio Bauz Olvera (Escuela Superio Politécnica del Litoral)

RESUMEN. En meta-análisis para pruebas de precisión diagnósticas (DTA) se utilizan medidas resumen como la sensibilidad, especificidad y odds ratio. Sin embargo, estas medidas pueden no ser adecuadas para integrar estudios con baja prevalencia, por tal motivo se hace necesario la modelización estadística basada en los verdaderos positivos y falsos positivos. En este contexto existen varios métodos estadísticos, El primero de ellos es un modelo de efectos aleatorios bivariante [Reitsma, 2005], parte de la asunción de que los logit de sensibilidad y especificidad siguen una distribución normal bivariante, El segundo, se refiere al enfoque HSROC o modelo jerárquico [Rutter CM and Gatsonis C., 2001], es similar al bivariante, con la particularidad de que modeliza de manera directa la relación sensibilidad y especificidad a través de puntos de corte. Utilizando simulaciones, investigamos el rendimiento de los modelos jerárquicos, variando sus parámetros e hiperparámetros y proponiendo un mejor manejo de la variabilidad dentro y entre estudios. Los resultados de los datos simulados se analizan de acuerdo al criterio de ajuste de los modelos, estimaciones de sus parámetros.

ESTADÍSTICA ESPACIAL

Moderador: **David Alejandro Ozuna Santiago** (Universidad Autónoma de Guerrero)

Modelos de regresión Waring Zero truncado

Luisa Isabel Rivas Calabrán (Universidad de Concepción)

RESUMEN. En este trabajo se presenta el modelo de regresión Waring Zero Truncado (ZTIW) para datos de conteo, obtenido a partir de la distribución del mismo nombre, y considerado una extensión del modelo de regresión Waring. Se obtienen los estimadores máximo verosímil, se presentan algunas propiedades, expresiones para el análisis de influencia. Los resultados son aplicados a un conjunto de datos reales y comparado con otros modelos plausibles a ajustar para dichos datos.

Construcción de un coeficiente de concordancia para datos espaciales

Javier Ignacio Pérez Ojeda (Universidad Técnica Federico Santa María)

RESUMEN. En este trabajo se define un coeficiente de concordancia para un proceso espacial bivariado de segundo orden estacionario. Se aborda el problema siguiendo el trabajo existente, enfocado ahora en un contexto espacial, considerando un coeficiente dependiente de la distancia fija entre dos localizaciones, de manera de comparar dos procesos espaciales con respecto a la línea de 45 grados con respecto al origen. Se presenta el coeficiente de concordancia espacial para estructuras de covarianza bivariada parámetricas, en particular para las covarianzas de Matérn y Wendland-Gneiting. Se establece la normalidad asintótica de una versión muestral del coeficiente de concordancia, en un escenario de dominio creciente, para ambas estructuras de covarianza, y además se realizan simulaciones computacionales para la estimación de parámetros del coeficiente vía máxima verosimilitud. Finalmente se presenta una aplicación para medir la concordancia espacial de imágenes forestales, tomadas por dos cámaras fotográficas en la misma ubicación (Harvard Forest), en donde se considerarán coeficientes de concordancia global, calculados a partir de concordancias medidas en sub-imágenes.

ENSEÑANZA Y DIVULGACIÓN DE LA ESTADÍSTICA

Moderador: **Vanessa Edith Figueroa** (Universidad Nacional De Catamarca)

Adaptación y validación del instrumento para medición de razonamiento estadístico SRA al contexto estudiantil chileno

Maria Paz Casanova Laudien (Universidad de Concepción)

RESUMEN. El razonamiento estadístico se define como la forma en que las personas razonan con ideas estadísticas y le dan sentido a la información estadística, habilidad esencial a desarrollar en estudiantes de carreras vinculadas a la investigación y mayormente a la estadística y matemática aplicada.

Garfield (1991) construyó un instrumento en inglés para medir razonamiento estadístico: Statistical Reasoning Assessment SRA.

En Chile, la estadística ha sido incorporada transversalmente al currículo de matemática en enseñanza básica y media (programas de matemáticas para todos los niños entre 6 y 17 años); sin embargo, esta decisión política no incluyó modificación de planes de estudio de las carreras de pedagogía en matemáticas, ni obligatoriedad de participar de cursos de actualización para profesores actualmente en ejercicio en el país. Esto lleva a cuestionar el nivel de razonamiento estadístico de los estudiantes que ingresan a la universidad.

Este trabajo presenta:

- Traducción, adaptación y validación en el contexto chileno de SRA.
- Propuesta de instrumento para ser utilizado en el medio educacional chileno y proveer estimación basal del razonamiento estadístico en escolares.
- Evaluación del efecto de la formación estadística universitaria sobre el razonamiento estadístico basal, mediante aplicación de SRA a alumnos universitarios de diferentes niveles y a docentes.

Identificación de los estilos de aprendizaje de los alumnos de bioestadística de las carreras de biología

Vanessa Edith Figueroa (Universidad Nacional de Catamarca)

RESUMEN. Los estilos de aprendizaje constituyen en un amplio campo de estudio y han llevado a considerar las diferencias personales de los alumnos en el proceso de aprendizaje, las que se relacionan con sus resultados académicos. En términos generales, los estilos de aprendizaje se definen como los rasgos cognitivos, afectivos y fisiológicos que sirven como indicadores relativamente estables de

cómo los estudiantes perciben, interaccionan y responden a sus ambientes de aprendizaje. Por ello, el conocimiento del estilo predominante en los alumnos se considera una herramienta docente relevante, permitiendo adaptar las estrategias de enseñanza a ellos y lograr así elevar el rendimiento académico del alumnado. El presente trabajo tiene como objetivo principal identificar los estilos de aprendizaje predominantes de los alumnos de la asignatura Bioestadística de las carreras de Profesorado y Licenciatura en Biología de la FaCEN- UNCa. y su vinculación con variables sociodemográficas. La investigación se enmarca en cuantitativo, de enfoque transversal, pues se analizaron los estilos de aprendizaje de los alumnos de segundo y tercer año de las carreras mencionadas, en los años 2017 y 2018, a partir de una encuesta para detectar los estilos de aprendizajes predominantes. Del análisis estadístico se observa que predominan los estilos Visual e Introvertido. En función de los resultados logrados se presenta una propuesta de estrategias de enseñanza de Bioestadística que responda a sus ambientes de aprendizaje. Se espera contribuir con esta propuesta a mejorar el rendimiento académico de los alumnos en Bioestadística.

Argumentaciones no convencionales en las pruebas de hipótesis

Hugo Granchetti Valdovinos (Universidad de Buenos Aires)

RESUMEN. En la actualidad, las Ciencias de la Salud incluyen a la estadística como herramienta de análisis y validación de resultados, por lo que su abordaje resulta de importancia en la formación de futuros profesionales en dicha área. Por lo tanto, el diagnóstico periódico de los aprendizajes de concepciones estadísticas fundamentales es esencial para la revisión constante de las estrategias de su enseñanza. En los últimos años se han evidenciado las dificultades que los alumnos presentan en la conceptualización del aprendizaje del nivel de significación y del nivel justo de significación, en pruebas de hipótesis. Dichas dificultades se manifiestan en la utilización de los modelos estadísticos sin la comprensión profunda de los conceptos subyacentes de probabilidad. Estos obstáculos se deben a diversas causas, entre ellas, la falta de articulación entre los contenidos teóricos y las prácticas asociadas, el uso mecánico de recursos informáticos en las propuestas didácticas, y la falta de experticia estadística de una parte del cuerpo docente. En este trabajo se propone un análisis sistemático cualitativo, desde una perspectiva socio-epistemológica, de las argumentaciones que aparecen en los exámenes finales de la asignatura Bioestadística, de alumnos de las carreras de Farmacia y Bioquímica de la Universidad de Buenos Aires. Para ello, se tomó una muestra de exámenes finales correspondientes a todas las fechas de evaluación desde el año 2015. Este análisis tiene como objetivo identificar argumentaciones no convencionales en las justificaciones escritas, en los ejercicios relacionados con las pruebas de hipótesis, con el fin de repensar la enseñanza de las ideas básicas de probabilidad, a lo largo de la asignatura.

ESTADÍSTICA Y CIENCIAS SOCIALES

Moderador: **Luz Judith Rodriguez Esparza** (Universidad Autónoma Chapingo)

Estimación de la prevalencia y factores de riesgo asociados a la violencia sexual contra las mujeres en zonas de conflicto armado en Colombia

Katerina Tinjaca Uriza (Universidad de la Salle)

RESUMEN. El conflicto armado en Colombia cuenta con antecedentes desde 1960 y aunque en la actualidad el mismo se mantiene, en septiembre del 2016 se firmó el acuerdo de paz entre el Gobierno y el grupo guerrillero más grande, las Fuerzas Armadas Revolucionarias de Colombia (FARC), terminando con más de 50 años de guerra. Esto dejó unas secuelas, que se evidencian principalmente en las zonas donde más se presentaron las confrontaciones, es por ello que a partir del acuerdo, se crea el Sistema Integral de Verdad, Justicia, Reparación y no Repetición, donde uno de sus componentes busca determinar la asociación entre la violencia sexual con el conflicto armado y cómo afectó a las mujeres.

particularmente para la problemática de la violencia sexual en contra de las mujeres, las estadísticas oficiales no dan cuenta de la realidad de este fenómeno por ser una violencia invisible tal como lo afirma Galtung (1998, 2003). Por ello, dentro de la campaña *violaciones y otras violencias: Saquen mi cuerpo de la guerra*, 14 organizaciones se unieron para diseñar una encuesta que permitiera levantar información de utilidad para la propuesta de políticas públicas que ayuden a disminuir el riesgo de ser víctima, den orientaciones sobre la reparación de las mismas y permita estimar la prevalencia de mujeres que sufrieron algún tipo de violencia sexual en zonas de conflicto armado en Colombia durante el periodo de los años 2010 a 2015. En conclusión, usando un análisis clúster se encontraron 3 grupos que permitieron caracterizar el fenómeno de violencia y a partir de un modelo logístico se identificaron 9 factores asociados a la violencia sexual contra las mujeres en zonas de conflicto armado

Una crítica a los modelos de panel cruzados para la estimación del efecto recíproco entre felicidad y desempeño laboral

Francisco Antonio Alfaro Medina (Universidad Tecnica Federico Santa Maria)

RESUMEN. Los modelos de panel cruzado (CLPM) han demostrado ser un procedimiento poco útil en el análisis de datos longitudinales. En particular, la diferencia entre las correlaciones cruzadas no forma una base sólida para la inferencia causal. Estos modelos carecen de una teoría sólida en aspectos tales como: teoría del cambio, selectividad de la muestra, efectos interindividuales e intra-individuales, entre otros. En este trabajo discutimos sobre los distintos modelos de ecuaciones estructurales

que incluyen relaciones cruzadas para abordar el problema de los efectos interindividuales e intraindividuales. En base a lo anterior, derivamos las ventajas y las limitaciones de cada uno de los modelos para establecer las relaciones de influencia entre la felicidad y el desempeño laboral para un conjunto de datos de tamaño 1000 correspondiente a trabajadores chilenos/as de múltiples organizaciones.

A Bayesian analysis of the impact of the freedom of the press in economic uncertainty

Rodrigo dos Santos Targino (Fundação Getulio Vargas (FGV))

RESUMEN. We show that variations on the level of the freedom of the press can have a substantial impact on different measurements of economic uncertainty. Using Freedom House's annual report to capture how the news media can freely operate in twelve countries, we investigate how changes in the freedom of the press impact financial market volatility and the economic policy uncertainty index. Using a Bayesian Vector Autoregressive Model (BVAR) with mixed frequencies, we provide the analysis for 8 of the OECD countries (Australia, Canada, France, Germany, Great Britain, Korea, Japan, and the United States), and the BRICs (Brazil, Russia, India, and China). As a by-product we use monthly financial data to provide an estimate of the monthly series for the FOTP index, which is released only once a year.

ESTADÍSTICA Y MEDIO AMBIENTE

Moderador: **Claudia Patricia Molinari** (Facultad de Farmacia y Bioquímica, Universidad de Buenos Aires)

Logit-linear modelling of Poisson point process: A bayesian option for presence-only type data

Guido Alberti Moreira (Universidade Federal do Rio de Janeiro)

RESUMEN. Species distribution models (SDMs) are extremely useful for determining preferences and habitats for different species. Appropriate estimation of species distribution depends on the adequate random sampling scheme which isn't always available. Instead, data is frequently composed of georeferenced locations where the species has been observed, which is commonly referred to as presence-only (PO) data.

The statistical modelling of PO type data through Inhomogeneous Poisson Processes (IPP) was proposed by Fithian and Hastie (2013). As has already been noted (Fithian et al, 2015), PO type data presents bias in its sampling pattern, which must be addressed. A natural way to model this bias under IPP is through thinning of the process, which is easily performed using pertinent covariates and a logistic function.

A different model for the intensity is proposed. It maintains the already established flexibility, while adding extra flexibility in the choice of covariates. Therefore, it is possible to have correlated and even the same covariates in the intensity and thinning components of the model. This is shown through simulated results. Additionally, it provides computational advantages for handling integrals that appear in the likelihood without resorting to approximations.

Finally, all the inference is based on the Bayes paradigm, which provides a way to consider the uncertainty inherent to the unknown components in the analysis.

PROBABILIDAD APLICADA Y PROCESOS ESTOCÁSTICOS

Moderador: **Alian Li Martín** (Universidad Autónoma de Guerrero)

Estimación de parámetros en procesos de Lévy t-Student mediante algoritmos de libre verosimilitud bayesiana

Manuel Jesús Pereira Barahona (Universidad del Bío-Bío)

RESUMEN. Al analizar conjuntos de datos temporales a partir de una variable continua, un proceso de difusión puede ser adecuado bajo ciertas condiciones de la distribución de los incrementos. Estamos interesados en procesos que presentan, por un lado, asimetría distribucional y por otro colas pesadas. En este trabajo se analizamos el método ABC-GIBBS que es libre de verosimilitud, para desarrollar inferencias en un proceso con incrementos independientes de Levy y distribución marginal t-Student. El Algoritmo ABC-GIBBS fue propuesto por Turner y Zandt (2012) en un ámbito de modelos de Psicometría, procedimiento que se aplica a modelos donde la función de verosimilitud es analíticamente tratable o computacionalmente prohibitiva que de alguna manera trata de eludir la evaluación explícita de la misma. La aplicación se realiza en datos del Índice de Precios Selectivo de Acciones (IPSA) desde el periodo 1-02-2001 hasta 28-08-2017.

El problema de la secretaria en procesos de decisión de Markov

Carmelo Hernandez Martinez (Universidad Juárez Autónoma de Tabasco)

RESUMEN. El Problema de la Secretaria en Procesos de Decisión de Markov Carmelo Hernández Martínez, Dr. Daniel Cruz-Suárez Universidad Juárez Autónoma de Tabasco División Académica de Ciencias Básicas

Los procesos de decisión de Markov (PDMs) dependen de una sucesión de decisiones (llamada Política) aplicadas en cada tiempo. Para evaluar la calidad de las políticas se cuenta, además, con cierto Criterio de Rendimiento definido en términos de un costo o recompensa por etapa. Entonces el Problema Básico de los PDMs (llamado Problema de Control Óptimo), consiste en optimizar el criterio de rendimiento sobre el conjunto de las políticas. A la política que optimiza el criterio de rendimiento se le llama política óptima. Un modelo en esta área es el problema de la secretaria, un modelo general aplicable en ramas como la ingeniería, economía y finanzas, entre otras, en el cual se busca optimizar cierto criterio de rendimiento. En el presente trabajo se analiza el problema de la secretaria, enfocado al ámbito financiero, específicamente en la optimización de un ejercicio de una opción americana, donde se busca que una vez adquirido un activo se emplee la opción de venta en el mejor momento antes de la fecha de vencimiento, en función del precio del mercado. Las opciones sobre acciones negociadas en mercados organizados suelen ser de tipo americano.

ESTADÍSTICA BAYESIANA

Moderador: **Luis Alberto Gutiérrez Inostroza** (Pontificia Universidad Católica de Chile)

A Bayesian mixture model for clustering circular data

Carlos Erwin Rodríguez Hernández-Vela (Universidad Autónoma Metropolitana, Iztapalapa)

RESUMEN. Clustering complex circular phenomena is a common problem in different scientific disciplines. For instance, the clustering of both departure directions and patterns of animals in the wild, the classification of changes in the timing of seasonal events into a changing climate frame and the grouping of observed X-ray positions from patients with radiation therapy represent challenging opportunities for different areas of research. In this paper, we propose a clustering and classification methodology for circular data under a Bayesian mixture modeling framework. The mixture model is defined assuming the number of components is finite, but unknown, and each component follows a projected normal distribution. Model selection is carried out by jointly making inferences about the parameters in the mixture model and the number of components, choosing the model with the highest posterior probability. A deterministic relabeling strategy is used to recover identifiability for the components in the chosen model. Estimates of both the posterior classification probabilities and the scaled densities are approximated via the relabeled MCMC output. Our proposal is illustrated using simulated and real data sets. The results suggest that the procedure is robust to a wide range of distributions for circular data.

Un modelo bayesiano no-paramétrico para datos circulares restringidos al primer cuadrante del plano cartesiano

Gabriel Núñez Antonio (Universidad Autónoma Metropolitana, unidad Iztapalapa)

RESUMEN. En el análisis de fenómenos reales existen variables direccionales que por su naturaleza se ven definidas solo en ciertos subconjuntos del círculo unitario. Por ejemplo, cuando se trabajan con datos *axiales*, el rango de posibles valores de las variables circulares asociadas resulta ser el intervalo $(0, \pi]$. Así, desde el punto de vista metodológico es importante contar con distribuciones de probabilidad definidas en subconjuntos acotados sobre el círculo unitario. Este trabajo pretende contribuir a la propuesta de modelos para describir variables circulares que no están definidas en todo el círculo unitario. Específicamente, en este trabajo se presenta un modelo bayesiano no-paramétrico para describir datos circulares multimodales restringidos al intervalo $(0, \pi/2]$. Se muestra la manera de llevar a cabo inferencias vía procedimientos MCMC. La metodología propuesta se ilustra analizando datos simulados.

ESTADÍSTICA Y SALUD PÚBLICA

Moderador: **Edgar Alejandro Guerrero Arroyo** (Universidad de Guadalajara)

Exposure to ambient PM2.5 concentrations and cognitive function among older Mexican adults

Aaron Salinas Rodriguez (Instituto Nacional de Salud Pública)

RESUMEN. Background: Recent epidemiological research has shown that exposure to fine particulate pollution (PM2.5) is associated with a reduction in cognitive function in older adults. However, primary evidence comes from high-income countries, and no specific studies have been conducted in low and middle-income countries where higher air pollution levels exist.

Objectives: To estimate the association between the exposure to PM2.5 and cognitive function in a nationally representative sample of older Mexican adults and the associated effect modifiers.

Methods: Data for this study were taken from the National Survey of Health and Nutrition in Mexico carried out in 2012. A total of 7,986 older adults composed the analytical sample. Cognitive function was assessed using two tests: semantic verbal fluency and three-word memory. The annual concentration of PM2.5 was calculated using satellite data. Association between exposure to PM2.5 and cognitive function was estimated using two-level logistic and linear regression models.

Results: In adjusted multilevel regression models, each 10 $\mu\text{g}/\text{m}^3$ increase in ambient PM2.5 raised the odds of a poorer cognitive function using the three-word memory test ($OR=1.37$, 95 % CI: 1.08, 1.74), and reduced the number of valid animal named in the verbal fluency test ($=-0.72$, 95 % CI: -1.05, -0.40). Stratified analyses did not yield any significant modification effects of age, sex, indoor pollution, urban/rural dwelling, education, smoking and other factors.

Conclusions: This study supports an association between exposure to PM2.5 concentrations and cognitive function in older adults. This is particularly relevant to low- and middle-income countries, which are marked by a rapid growth of their aging population and high levels of air pollution.

Modelos de regresión binaria: Aplicaciones para cáncer cervicouterino en una clínica de atención de la ciudad de Durango, Dgo. México

Edgar Felipe Lares Bayona (Instituto de Investigación Científica de la UJED)

RESUMEN. Objetivo. Aplicar y comparar modelos estadísticos de regresión binaria para Cáncer Cervicouterino (CaCu) en una clínica de atención de la ciudad de Durango, Dgo. México. Metodología. El método de investigación es deductivo, inductivo y de análisis. Es un diseño de estudio descriptivo para el análisis de regresión binaria. Se analizó una base de datos sobre resultados de papanicolau de una clínica de la ciudad de Durango para los años 2011 a 2014 con un total de 4939 pacientes, de

los cuales sólo 30 son positivas para CaCu. A este tipo de datos se les llama eventos raros. Resultados. Se estudiaron modelos de regresión binaria con funciones de enlace simétricas y asimétricas, un modelo de regresión logística para eventos raros y un modelo de regresión binaria Bayesiano semi-paramétrico. Los modelos con enlaces simétricos y asimétricos muestran resultados similares en la población estudiada. El modelo logístico corregido para eventos raros obtuvo una mejor evaluación diagnóstica del modelo, estimando una mayor cantidad de verdaderos positivos. El modelo Bayesiano obtuvo coeficientes más acordes al contexto para CaCu. Un análisis posterior de submuestras en un proceso iterativo de aleatorización, identificó a covariables que se pueden seguir manteniendo dentro del modelo y a otras covariables que no mantienen resultados acordes al contexto, son considerados como sesgos no propios del análisis estadístico.

APRENDIZAJE DE MÁQUINA

Moderador: **Mauricio Castro Cepero** (Pontificia Universidad Católica de Chile)

Módulo de minería de textos para el análisis multivariante textual en R

Luis Jesus Lozano Duran (Universidad Veracruzana)

RESUMEN. La minería de textos parte de la Estadística y de las Ciencias de la Computación, la cual se encarga de la extracción, recuperación, gestión y análisis de la información, dada su complejidad escasamente se encuentran herramientas tecnológicas que contengan la sustracción y manipulación de información. Analizar grandes volúmenes de datos se ha vuelto una necesidad debido a la creciente información de medios digitales e impresos, sin embargo, la extracción de esta toma tiempo en el proceso de selección y sintetizado, ya que gestionar los datos consta del limpiado de caracteres, ordenamiento lexical, eliminación de espacios, estructuración de la base y creación de corpus o tokens. La implementación del análisis textual suele tener menos relevancia en estudios de investigación aplicada sobre contenidos textuales, debido a la poca implementación de técnicas estadísticas que permitan la identificación de patrones de comportamiento, clasificación lexical y estructuración taxonómica de las palabras frecuenciales. Para realizar un análisis textual se requiere de estructurar los textos en grupos que tienen cualquier dimensión o longitud y se detecta el idioma en el que se encuentra el formato del archivo. Es por ello, que el presente trabajo de investigación se enmarca en la proposición de un Módulo de Minería de Textos Multivariantes (MTMM) que permita a través de técnicas multivariantes textuales mostrar patrones de comportamientos de corpus de información. Este módulo contiene diversas secciones para la recuperación, procesamiento y análisis, facilitando así la evaluación de contenidos textuales en diversas áreas de aplicación.

Un enfoque basado en muestreo local para la solución del problema de Máquinas de Vectores de Soporte

Roberto Bárcenas Curtis (Centro de Investigación en Matemáticas A.C.)

RESUMEN. El objetivo de este trabajo es desarrollar una metodología de clasificación para la solución del problema de Máquinas de Vectores de Soporte (SVM) basada en un muestreo local, considerando un esquema que proporcione información en la búsqueda de observaciones cercanas a los vectores de soporte.

La idea de muestreo local en este contexto consiste en utilizar las soluciones de varias submuestras pequeñas del conjunto de entrenamiento y enriquecer el procedimiento con observaciones de la muestra completa, mediante un muestreo de mayor intensidad y bajo ciertos criterios de búsqueda en las regiones en donde se detecta la existencia de mayor número de vectores de soporte.

Adicional a la descripción algorítmica, se plantean un par de resultados para sustentar sus propiedades teóricas. El primer resultado va en el sentido de considerar la cercanía de observaciones de una submuestra a los vectores de soporte. En el segundo resultado se establece una cota para cuantificar la aproximación entre el problema resuelto con la totalidad de observaciones y la solución basada en submuestreo.

El desempeño de este enfoque es evaluado en la práctica utilizando conjuntos de datos provenientes del sitio UCI Machine Learning Repository de la Universidad de California en Irvine. Los ejemplos son variados tanto en la cantidad de observaciones como en cuanto a su dimensión. Los resultados preliminares muestran que se tiene una buena aproximación al error de entrenamiento con el empleo de una menor cantidad de vectores de soporte y en términos de la reducción del tiempo, los resultados también son favorables.

ANÁLISIS DE DATOS

Moderador: **Iván Pacheco Soto** (Instituto Tecnológico y de Estudios Superiores de Monterrey)

Análisis de consonancia aplicado a la evaluación de atributos por parte de un panel de expertos en el contexto de una elección rectoral

Claudio Rafael López Castro (Universidad Veracruzana)

RESUMEN. El Análisis de Consonancia es una poderosa herramienta en el área de los métodos multivariantes, puede ser utilizada para evaluar la consistencia entre una serie de juicios o percepciones que emite un panel de expertos sobre un tema o área determinados, o bien sobre un conjunto de atributos de personas que aspiran a un determinado cargo o puesto de dirección. Lo que da pauta enormes posibilidades de aplicación en estudios de opinión. A partir de una matriz de datos de orden ($N \times M \times K$), donde K es el número de evaluadores expertos, N el número de personas participantes y M los atributos a evaluar por parte de los expertos. El método realiza la descomposición de eigenvalores de la matriz de datos, evaluada a través de el Análisis de Componentes Principales. Usamos el programa CONS, desarrollado en la plataforma R. Por otra parte, las formas de selección de las máximas autoridades en universidades públicas, en particular de los rectores, muestran diferencias significativas en México y el mundo. La legislación de las universidades públicas en México presentan cuatro modelos básicos (López, 2010, López, et. al., 2011a, 2011b, López, 2013; Ordorika, et. al., 2011, Ordorika, 2014, 2015). Los modelos o métodos de selección son: 1. Elección por voto directo de la comunidad universitaria; 2. Elección por órgano colegiado-representativo (consejos o asambleas universitarias); 3. Designación por juntas de gobierno o directivas; y, 4. Nombramiento por órgano externo (gubernamental y/o empresarial). En la Universidad Veracruzana, en México se elige el rector a través de una junta de gobierno, compuesta por 9 miembros (académicos notables), 5 pertenecen a la comunidad UV y 4 son académicos externos a ésta. Se presentan los resultados del Análisis de Consonancia en la evaluación de aspirantes a Rector, en base a la aplicación de cuestionario a cada miembro de la junta de gobierno, en la fase final del proceso de elección rectoral.

Desarrollo de Estadísticas para Medidas de Conceptos Culturales Medidos por Cambios Lingüísticos

Mirtha Haydee Pari Ruiz (Universidad de Valparaíso)

RESUMEN. La presente investigación hace un estudio comparativo de frecuencia de palabras en dos diferentes corpus del Nican Mopohua (relato Náhuatl-Méjico). Empleamos el corpus de Lazo (1649) y Rojas (1978). Se aplica el método probabilístico de Kullback Leibler (KL) divergence para cada palabra analizada y se demuestra que recibimos diferentes divergencias. Por ejemplo analizando la palabra “xochitl” (flores) de un corpus a otro recibimos valores de 0.618718998 y 0.272281705 . Lo que significa

que el lenguaje (palabra) no presenta un comportamiento para medidas simétricas ya que presenta variaciones por el tiempo por lo que su estudio se hace más propicio en el espacio topológico (Stehlik y Pari Ruiz 2018). Está presencia más asimétrica en las distancias entre palabras se basan más en lo topológico y mereológico y en las nociones de conexión, parte, interior y complemento ya que la semántica de la expresión del lenguaje natural concierne a ubicación y posición relativa (Stehlík, 2016)

Social automation, twitter bots and human rights

Pablo Suarez Serrato (Instituto Matemáticas UNAM)

RESUMEN. As social media use has proliferated, so has the appeal for powerful actors to abuse platforms and bend public perception. In 3 case studies I will show, we will see how Twitter socialbots have acted by preventing/promoting information access, and creating a positive fog to cloak protests. We will focus on certain recent events in Mexico and Honduras. In relation to the disappearance of the 43 students from Ayotzinapa, Guerrero and ensuing protests online using #YaMeCanse; on the promotion of access to a human rights abuse report about a raid in a ranch in #Tanhuate, Michoacán; on the Honduran presidential election at the end of 2017; and on harrassment of journalists and impunity prosecutors in Guatemala. As the tweet corpora are—almost completely—in Spanish, the available tools have to be somewhat adapted for our analysis.

ANÁLISIS DE SUPERVIVENCIA

Moderador: **Nora Beatriz Cadenas** (Dirección Nacional de Observatorio Argentino de Drogas y Universidad Nacional de Tres de Febrero)

A flexible joint model for time-to-event and a sequence of longitudinal biomarker data, a full dynamic hierarchical approach

Pamela Massiel Chiroque Solano (UFJR)

RESUMEN. This work proposes a class of full dynamic hierarchical joint model that relates the simultaneous behaviour of a sequence of longitudinal measurements and event times. To allow for greater flexibility, key components of the model are treated dynamically. For the subject-specific longitudinal evolutions we use a polynomial generalized dynamic hierarchical approach and the semi-parametric approach the baseline hazard function is assumed piecewise constant. The estimation is done under the Bayesian paradigm using MCMC methods, we analyze flexible classes of prior distributions for modeling the failure rate was analyzed. To compare models measures are developed based on the decomposition of the DIC and LPML allowing determines the relevance and contribution of the longitudinal trajectory on the model; the performance between the proposed model and standard models is compared. The methodological novelties are illustrated with a simulation study and two applications to real data to examine performance of the proposed model. Furthermore, we use Bayesian hypothesis tests to understand unexplained variability; the results show the importance of the use of our propose model.

Análisis de supervivencia con medidas aleatorias compuestas

Alan Riva Palacio Cohen (University of Kent)

RESUMEN. Análisis de supervivencia es una de las primeras áreas donde se realizaron aplicaciones de la estadística Bayesiana no paramétrica. Por otro lado, recientemente ha habido gran interés en modelos multivariados que permitan estructuras de dependencia. Proponemos un modelo de supervivencia para información basada en muestras múltiples y un modelo de regresión en supervivencia. Nuestras construcciones se basan el uso de vectores de medidas completamente aleatorias dependientes. Mostraremos resultados explícitos acerca de las distribuciones posteriores asociadas. Ejemplos con medidas aleatorias compuestas, una familia flexible de vectores de medidas aleatorias, son presentados.

Procesos de Poisson espaciales para el inicio de grietas en el comportamiento a fatiga de materiales metálicos

Marco Scavino (Universidad de la República - Facultad de Ciencias Económicas y de Administración)

RESUMEN. En este trabajo proponemos un modelo estocástico para estimar la probabilidad de ocurrencia del inicio de grietas en la superficie de especímenes metálicos en problemas de fatiga, cuya aplicación pueda extenderse a una clase general de geometrías. Nuestro objetivo principal es estimar la probabilidad de supervivencia de cualquier componente mecánica empleando la información que proviene de datos de experimentos de fatiga en muestras hechas del mismo material. El modelo estocástico propuesto se basa en procesos espaciales de Poisson cuya función de intensidad combina las curvas S-N con una función de tensión espacial, llamada tensión efectiva promedio, que se calcula después de resolver numéricamente las ecuaciones de elasticidad lineal en el dominio del espécimen, utilizando el método de los elementos finitos. La tensión efectiva promedio, parametrizada por una cantidad que depende del material considerado, transforma el tensor tensión en un campo escalar definido sobre la superficie del espécimen. Para calibrar los parámetros del modelo, tanto para los conjuntos de datos individuales como para su combinación, hemos empleado datos que refieren a diversos registros de experimentos de fatiga en muestras de hojas de aleaciones de aluminio 75S-T6 con y sin entalladuras. Una vez obtenida la forma exacta de la función de log-verosimilitud del modelo, hemos aplicado el modelo, desde una perspectiva clásica y bayesiana, estimando la función de probabilidad de supervivencia para cada tipo de geometría del espécimen, sometido a una determinada configuración experimental de fatiga. Finalmente, hemos comparado las estimaciones a posteriori de las funciones de supervivencia para los datos combinados con las estimaciones de máxima verosimilitud de referencia para los conjuntos de datos individuales.

ANÁLISIS MULTIVARIADO

Moderador: **Carlos Díaz Avalos** (Universidad Nacional Autónoma de México)

Fuzzy HJ-biplot como herramienta para la minería de datos

Ana Belém Nieto Librero (Universidad de Salamanca)

RESUMEN. En diversas aplicaciones de la Minería de Datos es necesario disponer de técnicas que combinen la clasificación de objetos y por otro lado la búsqueda de patrones mediante técnicas de reducción de la dimensión. Recientemente, se ha propuesto un método denominado Clustering HJ-Biplot en el que se pretende dar solución al doble problema de clasificación y representación de los resultados en baja dimensión. Sin embargo, al utilizar el k-means como técnica de clasificación, presenta el problema del agrupamiento exclusivo. En este tipo de métodos de clasificación se considera que cada elemento se puede agrupar inequívocamente con los elementos de su cluster y que, por lo tanto, no se asemeja al resto de los elementos. Para solventar esta deficiencia aparecen técnicas basadas en la lógica difusa como es el fuzzy c-means en el que para cada observación se calcula el grado de pertenencia a cada cluster.

En este trabajo se propone el método Fuzzy HJ-Biplot el cual tiene por objetivo encontrar la mejor clasificación de los objetos en un espacio de dimensión reducida utilizando la lógica difusa y el HJ-Biplot.

Métodos de clasificación para el diagnóstico automático de la enfermedad de Parkinson a través de biomarcadores de grabaciones de voz

Ruth Fuentes García (UNAM)

RESUMEN. Existen propuestas que han estudiado diferentes métodos de clasificación para distinguir a pacientes sanos de pacientes con Enfermedad de Parkinson, considerando grabaciones de voz. Comúnmente las medidas repetidas por individuo (distintas grabaciones por individuo) se han promediado para obtener una sola medición por individuo. El interés de este trabajo evaluar los métodos de clasificación, considerando medidas repetidas de los sujetos (las distintas grabaciones de cada sujeto). Se busca reducir la dimensión de las variables, y utilizar una modelación adecuada de las variables, considerando la naturaleza multi-modal de las mismas. Lo anterior permite observar mejoras en las medidas de ajuste como especificidad y sensibilidad del diagnóstico.

Predicción espacial y temporal del valor de un inmueble

Leonardo Moreno Romero (Departamento de Métodos Cuantitativos, Udelar)

RESUMEN. El objetivo del trabajo es construir un modelo que permita efectuar de manera eficiente,

en términos predictivos, la estimación del valor contado de un determinado inmueble. Se tiene como insumo la información espacio-temporal sobre el valor contado de un conjunto de inmuebles y ciertas variables hedónicas intrínsecas a cada bien, por ejemplo, superficie, antigüedad, número de dormitorios. En tal sentido es conocido el valor de ciertas propiedades en diferentes fechas, entendiendo por valor contado aquel que es asignado por el tasador, donde cada propiedad puede haber sido tasada en diferentes momentos, problema denominado en la literatura como ventas repetidas. La información espacial permite la construcción de modelos autorregresivos espaciales donde el precio de un inmueble se encuentra correlacionado con el de sus vecinos. La información temporal es modelada mediante un modelo autorregresivo temporal, en este caso la dependencia entre una tasación y la siguiente disminuye en función del tiempo transcurrido entre una y otra. Las variables hedónicas se modelan mediante un modelo de regresión lineal dinámico, donde los coeficientes de la regresión son función de la ubicación del inmueble. A partir de la metodología de agregación de Stacking se busca un procedimiento que permita predicciones más precisas. Se evalúa la performance mediante el error cuadrático en una muestra de testeo.

VIERNES 5 DE OCTUBRE
12:30 A 14:00 HRS

PONENCIAS (CONTRIBUCIONES LIBRES)
AUDITORIO ANTONIO ALATORRE

ENSEÑANZA Y DIVULGACIÓN DE LA ESTADÍSTICA

Moderador: **María Luisa Zepeda Cruz (UNAM)**

Una aplicación web en R como auxiliar para el aprendizaje de la regresión lineal

María Luisa Zepeda Cruz (Facultad de Ciencias UNAM)

RESUMEN. El análisis de regresión lineal es una de las primeras herramientas estadísticas que se enseñan a los estudiantes, de estadística y en general de casi cualquier disciplina. Esta técnica permite entender algunos de los conceptos básicos de la modelación estadística y sienta las bases para el estudio de modelo más complejos. El propósito de la presentación es mostrar una aplicación web interactiva, escrita con el paquete Shiny de R, que pretende servir de herramienta auxiliar en el aprendizaje de la regresión lineal. La aplicación permite explorar de manera interactiva algunos de los temas relevantes en el análisis de regresión lineal: análisis exploratorio, ajuste del modelo, inferencias por intervalos, contraste de hipótesis, validación de supuestos y selección de variables. El módulo de validación de supuestos es el más importante ya que permite entender el efecto que tienen en las inferencias posibles desviaciones a los supuestos del modelo.

VIERNES 5 DE OCTUBRE
16:00 A 17:00 HRS

PONENCIAS (CONTRIBUCIONES LIBRES)
AUDITORIO ANTONIO RODRÍGUEZ

ESTADÍSTICA Y SALUD PÚBLICA

Moderador: **Aaron Salinas Rodríguez** (Instituto Nacional de Salud Pública)

A Neural Network Approach in the Diagnosis of Patients with Depression

Edgar Alejandro Guerrero Arroyo (Universidad de Guadalajara)

RESUMEN. Depression is without a doubt a big issue of social health. Patients with this psychological disorder experience deep sadness, feel of hopeless and certain other symptoms that can lead to problems in crucial personal aspects such as family, social life, work, finance and even committing suicide. Therefore, being able of perform an accurate diagnosis of depression in patients it becomes crucial in order to establish a suitable treatment.

In the present talk we will show a Neural Network approach to help with the diagnosis of patients with depression based on their symptoms. Computational aspects and some results using real data will be discuss.

Estudio de simulación comparando el alpha de Cronbach y el alpha Ordinal

Sergio Eduardo Contreras Espinoza (Universidad del Bío-Bío)

RESUMEN. El objetivo principal de este estudio fue analizar las diferencias en la estimación de la confiabilidad de un instrumento usando los estimadores alpha de Cronbach y alpha Ordinal en distintos contextos simulados, estos fueron diferentes números de ítems, opciones de respuesta, tamaños muestrales y valores de confiabilidad, tanto para casos simétricos como asimétricos.

Para menos de 6 opciones de respuesta, alpha Ordinal es un mejor estimador que alpha de Cronbach en todos los casos simulados. Respecto del efecto de la asimetría de los datos, ésta perjudica la estimación de alpha de Cronbach, no teniendo un efecto tan grave en el caso de la estimación utilizando alpha Ordinal.

VIERNES 5 DE OCTUBRE
16:00 A 17:00 HRS

PONENCIAS (CONTRIBUCIONES LIBRES)
AUDITORIO ANTONIO ALATORRE

ESTADÍSTICA BAYESIANA

Moderador: **Carlos Erwin Rodríguez Hernández-Vela** (Universidad Autónoma Metropolitana, Iztapalapa)

Estimación bayesiana de la tendencia del voto en las elecciones presidenciales de México

Javier Santibáñez Cortés (Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas)

RESUMEN. En este trabajo se presenta un modelo jerárquico bayesiano para la estimación conjunta de p totales en una población finita, a partir de una muestra seleccionada con un diseño aleatorio estratificado. La estrategia de estimación consiste en utilizar la distribución predictiva para estimar la parte no muestreada de cada estrato, o subpoblación, para estimar los totales en cada estrato como el total en la muestra mas la predicción del total en el complemento de la muestra. El modelo se aplica a la estimación de la tendencia de voto en las elecciones presidenciales en México a partir de una muestra de casillas seleccionada con un diseño estratificado por distrito electoral federal. Se utiliza un algoritmo MCMC para simular observaciones de la distribución predictiva de las votaciones totales por partido. Los resultados obtenidos se comparan con los de otras estrategias de estimación utilizadas en algunos de los conteos rápidos que ha realizado el INE.

*Existencia y control de error de la distribución posterior en el problema inverso bayesiano
(Bayesian UQ) en espacios de Banach*

José Andrés Christen Gracia (CIMAT)

RESUMEN. Presentamos una generalización a espacios de Banach de resultados anteriores de los autores, en el que se hace uso de factores de Bayes para comparar la distribución posterior numérica con la teórica. Para esto, en el caso infinito dimensional de espacios de Banach se demuestra su existencia con condiciones más débiles que lo que la comunidad de UQ ha venido considerando. El resultado fundamental es una cota para el máximo error que el solver numérico puede tolerar, manteniendo la posterior numérica prácticamente indistinguible de la posterior teórica.

VIERNES 5 DE OCTUBRE
16:00 A 17:00 HRS

PONENCIAS (CONTRIBUCIONES LIBRES)
AUDITORIO MATUTE REMUS

ANÁLISIS DE DATOS

Moderador: **Roberto Bárcenas Curtis** (Centro de Investigación en Matemáticas)

Un nuevo método para determinar el nivel de traslape de dos distribuciones

Eduardo Gutiérrez Peña (Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas)

RESUMEN. En este trabajo proponemos un procedimiento sencillo para determinar el nivel de traslape de dos distribuciones. Este procedimiento es particularmente útil cuando las distribuciones son multivariadas y de una dimensión relativamente alta. Discutiremos posibles aplicaciones e ilustraremos el método con el análisis de un conjunto de datos ecológicos.

PRUEBAS DE HIPÓTESIS

Moderador: **Abelardo Montesinos López** (Universidad de Guadalajara)

Test de bondad de ajuste para la distribución Hermite bivariante

Pablo Andrés González Albornoz (Universidad del bio-bio)

RESUMEN. Los datos de conteo pueden aparecer bajo diferentes circunstancias. En un marco univariante, la distribución Poisson es la que con mayor frecuencia ha sido empleada para modelar tales datos. En la práctica, los datos de conteo bivariantes surgen en varias disciplinas diferentes y la distribución Poisson bivariante (DPB), siendo una generalización de la distribución Poisson. Un caso particular de la DPB es la distribución de hermite bivariante (DHB) juega un rol importante en modelarlos, siempre que dichos datos presenten una correlación no negativa. La distribución de Hermite es cero-inflada respecto a la distribución de Poisson, lo cual significa que la probabilidad del cero en la distribución de Hermite es siempre superior a la probabilidad del cero de la Poisson, cuando ambas distribuciones tienen la misma esperanza. También permite modelizar datos en los que aparece el fenómeno de sobredispersión, esto ocurre cuando la varianza es mayor que la esperanza, siendo esta sobredispersión moderada.

Contrastar la bondad de ajuste de las observaciones dadas con las hipótesis distribucionales asumidas es un aspecto crucial del análisis de datos. Por esta razón, en este trabajo estudiamos y proponemos un test de bondad de ajuste para la DHB. Este test se basara en la función generatriz de probabilidades, el cual sea consistente contra alternativas fijas.

Pruebas de bondad de ajuste para la distribución normal asimétrica basadas en transformaciones

Waldenia Cosmes Martínez (Colegio de Postgrados)

RESUMEN. En los últimos años se han desarrollado diversos métodos estadísticos que involucran datos provenientes de la familia de distribuciones normal asimétrica. De aquí la importancia de contar con pruebas de bondad de ajuste que permitan decidir si un conjunto de datos proviene de dicha familia.

Existen diversos trabajos que abordan este tema, la mayoría de estos, basados en estadísticos que miden la desviación entre la función de distribución de una población normal asimétrica y la función de distribución empírica. Otros utilizan las propiedades de la distribución normal asimétrica para construir estadísticos de prueba específicos. En el presente trabajo se aborda el caso en el que los valores de los parámetros son desconocidos. Se proponen dos procedimientos de prueba, el primero basado en una transformación a variables normales y el segundo a variables gamma. Para probar la hipótesis de normalidad sobre los datos transformados se utiliza la prueba de Shapiro-Wilk y para probar

la hipótesis de distribución gamma se utiliza Anderson-Darling. Adicionalmente se implementa la prueba de Anderson-Darling directamente sobre la muestra, pero utilizando bootstrap paramétrico para el cálculo de los valores críticos.

Se presentan resultados de estudios de simulación sobre el tamaño de las pruebas y su potencia frente a diversas distribuciones alternativas, las cuales fueron escogidas por sus características e importancia en la modelación de datos con asimetría. Además, se presenta una clasificación de dichas alternativas de acuerdo con la potencia de las pruebas propuestas y otras ya publicadas. Lo anterior con el fin de identificar qué pruebas son mejores en cada caso para distintos tamaños de muestra.

ESTADÍSTICA OFICIAL

Moderador: **Waldenia Cosmes Martínez** (Colegio de Postgraduados)

Intervalos de confianza en medicina con enfoque superpoblacional

Juan José Goyeneche Capeluto (Universidad de la República)

RESUMEN. En Medicina es común encontrarse con tasas relativamente pequeñas (tasa de mortalidad infantil, prevalencia de enfermedades raras, etc.), expresadas en tanto por mil o tanto por diez mil para las cuales se desean construir intervalos de confianza. En general se dispone de datos poblacionales, por lo cual más que una estimación se dispone del dato verdadero. Los autores proponen un enfoque superpoblacional para resolver el problema, en el concepto de superpoblación o superuniverso introducido por Stouffer en 1934, quien lo definió 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” (Silva Aycaguer, 1997)

Los valores cercanos a cero de las tasas, aunque haya una cantidad muy numerosa de datos, hacen que la aproximación usando el Teorema Central del Límite no sea lo más apropiado. Se comparan tres métodos de la inferencia clásica para construir los Intervalos de Confianza: método de Wald, o aproximación Normal, método “exacto” a partir de invertir una prueba de hipótesis y método del score, el cual presenta mejores resultados de cobertura. Se presenta asimismo un método bayesiano para construir intervalos de credibilidad que tiene porcentajes de cobertura y precisión similares a los métodos clásicos.

Los cuatro métodos se aplican a tasas de mortalidad infantil para los años 2004 a 2016 en Uruguay, para el país en su conjunto y para subregiones.

Small area estimation based on a two-fold nested error lognormal model

Georges Bucyibaruta Bucyibaruta (Centro de Investigación en Matemáticas)

RESUMEN. Estimation of non-linear small area population means under a two-fold log-linear unit-level model is proposed. Closed form expressions, based on bias correction, for an empirical Bayes (EB) predictor and approximation for the associated mean squared error (MSE) estimator, in the framework of Prasad and Rao (1990), are obtained. We showed by simulations that empirical Bayes predictors are well behaved in terms of Mean squared error. We also proposed a parametric bootstrap method for MSE estimation and studied its bias in simulations.

Estimación basada en modelos del ingreso promedio de los hogares en los municipios de México

María Luisa Zepeda Cruz (Facultad de Ciencias UNAM)

RESUMEN. La medición de los ingresos en los hogares es un insumo valioso para la evaluación de las políticas públicas y las acciones del gobierno para la disminución de la pobreza en México. El INEGI proporciona estimaciones confiables del ingreso en los hogares a nivel nacional y con ciertas desagregaciones, pero no hasta la escala de municipios. En este trabajo se presenta una propuesta de estimación del ingreso promedio en los hogares de los municipios de México a partir de variables socio-económicas y basada en un modelo lineal ajustado con información de la Encuesta Nacional de Ingresos y Gastos en los Hogares del INEGI y de la Encuesta intercensal 2015. Se presenta un análisis descriptivo los resultados obtenidos a escala municipal y un análisis comparativo a escala de entidades de los resultados con distintas estrategias de estimación. Se aplican técnicas de selección de variables y de validación cruzada para construir el mejor modelo. La ventaja del método propuesto radica en que las estimaciones se basan en un modelo por lo que, bajo ciertas condiciones, el diseño de muestreo con el que fue seleccionada la muestra es ignorable.

ESTADÍSTICA BAYESIANA

Moderador: **José Andrés Christen Gracia** (CIMAT)

Enfoque bayesiano para la estimación de parámetros en sistemas dinámicos de la enfermedad del Dengue.

Alian Li Martín (Universidad Autónoma de Guerrero)

RESUMEN. El Dengue es una enfermedad vírica transmitida por el mosquito hembra de la especie *Aedes Aegypti*, esta enfermedad surge donde se encuentre los tres factores: el vector, el virus y un huésped susceptible. Por esta razón se distribuye en países de África, Asia, Oceanía y América, en donde se encuentran las condiciones geográficas, demográficas y epidemiológicas propicias. La presencia en México de esta enfermedad se ha incrementado en los últimos 14 años, en 30 estados de la república. Esta epidemia se determina por su rápida extensión y larga duración, debido a esto se ha hecho necesario llevar a cabo varios estudios de la misma, para así entender mejor la epidemia y saber qué medidas tomar para su control. Un enfoque de estudio de esta enfermedad es el desarrollo de modelos matemáticos y estadísticos que describen la dinámica viral, los mecanismos implicados en la transmisión del patógeno entre el huésped y los vectores, los factores que influyen en la incidencia del Dengue, entre otros. En la literatura relacionada con la estimación de parámetros en sistemas dinámicos de la enfermedad del Dengue se hace común el uso de métodos numéricos para encontrar la solución del sistema de ecuaciones diferenciales y luego el uso de métodos de optimización para la estimación de parámetros. Trayendo esto consigo que en sistemas dinámicos donde se tengan gran cantidad de ecuaciones y parámetro, el cómputo se haga lento y costoso. Este trabajo de investigación pretende realizar un estudio comparativo de los métodos existentes para la estimación de parámetros de sistemas dinámicos y proponer un modelo bayesiano específico, en el caso del Dengue y las respectivas estimaciones de sus parámetros.

Towards a bayesian solution of an inverse gravimetry problem

Abel Palafox González (Universidad de Guadalajara)

RESUMEN. In Geophysical applications, it is important to know what type of material is present into the subsoil and how it is placed. The target is to identify if this material might be a water deposit, fracture, solid rock, etc. Geophysical exploration by means of drilling is highly costly. Thus, physical properties (such as magnetic, gravitational or electrical fields) are measured in the surface of a terrain in order to envisage the characteristics of the corresponding subsoil. In this work, we pose a Bayesian formulation for the inverse problem of estimating the density of a rectangular object inside of a terrain from measurements of the gravitational field at the surface. Numerical implementation and results with synthetic data are also discussed.

Un modelo de Markov oculto para el monitoreo de la progresión de la enfermedad de Parkinson

Lizbeth Naranjo Albarrán (Universidad Nacional Autónoma de México)

RESUMEN. Motivado por un estudio sobre el monitoreo de la progresión de la enfermedad de Parkinson basado en características extraídas de grabaciones de voz, se propone un modelo de Markov oculto no homogéneo con espacio de estados continuo. El modelo considera que la variable respuesta está sujeta a un error de medición, y que el comportamiento debería ser monótono no decreciente; además debido a las imperfecciones en la tecnología y a la variabilidad biológica se tienen características de voz replicadas no idénticas, lo que resulta en covariables con medidas repetidas correlacionadas. La estimación de los parámetros se realiza usando métodos de Estadística Bayesiana.

ESTADÍSTICA Y MEDIO AMBIENTE

Moderador: **Martha María Tellez Rojo Solis** (Instituto Nacional De Salud Publica)

El gasto de energía y las economías de escala en la generación de CO2 por los hogares: un análisis en las regiones de México

Sergio Colin Castillo (Universidad Autónoma de Coahuila)

RESUMEN. Los estudios acerca de la cantidad de emisiones de dióxido de carbono (CO2) que emiten los hogares son escasos, y para México prácticamente nulos. Intentando llenar este hueco en la investigación, este trabajo estudia los patrones de gasto en energía (gasolineras, diésel, gas natural, gas LP y electricidad) y sus correspondientes niveles de emisión de CO2 emitidos por los hogares de México. Considerando que de 1984 al 2016, la población pasó de 75.78 a 127.5 millones de habitantes, y los hogares crecieron de 15.03 a 33.46 millones, pero en tasas de crecimiento, la población decrece (2.13 al 1.31) y el número de hogares crece al 5.9 %, donde lo más destacable es la evolución de los hogares unipersonales (una persona) con tasas promedio del 11.7 %. Esta es una tendencia global que indica que los hogares se vuelven cada vez más pequeños. En efecto, para el periodo antes señalado, el tamaño medio de los hogares mexicanos pasó de 5.1 a 3.67 habitantes, y la proporción de hogares unipersonales con respecto al total de hogares pasó de 5.2 % al 10.7 %. Para ligar este crecimiento de la población, tamaño del hogar, y generación de CO2, se utiliza el concepto de economías de escala de CO2 por hogar: A medida que los hogares crecen en número de habitantes, su generación per cápita de CO2 disminuye. Pero más allá de medir como es esta pauta global en México, la cual no ha sido antes medida, se explora su relación con el ingreso, entre otras variables. Una medición como la aquí propuesta es relevante en política económica y ambiental, pues permite dilucidar si los esfuerzos para disminuir los niveles de CO2 son fructíferos, y además abre un abanico de opciones en política de mitigación, medidas que por el lado de la demanda pudieran ser efectivas. La base de datos utilizada es la ENIGH (70,311 observaciones), representativa por Estado y en los ámbitos rural y urbano, y con sus factores de expansión es posible hacer inferencia en los 33462,598 hogares de México, codificando con Stata15

*Modelación espacial de la infestación del acaro rojo de las palmas (*Raoiella Indica*) en Guerrero*

David Alejandro Ozuna Santiago (Universidad Autónoma de Guerrero)

RESUMEN. El estado de Guerrero es el primer productor de palma de cocotero, con una superficie actual entre 40,000 a 50,000 hectáreas dedicadas a dicho cultivo. Dado que la palma de cocotero tiene un impacto económico importante en varios sectores de la sociedad, es necesario entonces no descuidar aspectos fitosanitarios para asegurar producción y calidad. Una manera de hacerlo es

con acciones que controlen las plagas que lo afectan; una de ellas es el ácaro rojo, que ocupa el cuarto lugar en pérdidas de producción, de ahí la importancia de su control. Este trabajo, tiene como objetivo modelar la distribución espacial y estudiar la proliferación de la plaga del Acaro Rojo de las palmas en las zonas de costa grande, costa chica y Acapulco del estado de Guerrero; mediante diversas técnicas de interpolación para poder detectar estructuras de correlación y así identificar las zonas de mayor infestación. Esto es muy importante en la agricultura de precisión, pues permite una mejor dirección de los controles químicos, logrando así una reducción de los costos por el manejo del cultivo. La dirección de las medidas de control sobre zonas específicas disminuye la resistencia de la plaga, trae beneficios a la salud de las personas y no se contamina más al medio ambiente. El estudio fue realizado con datos proporcionados por SENASICA de los años 2017. Tomando en cuenta la metodología y la base de datos, este estudio es el primero realizado hasta el momento para el ácaro rojo de la palma de cocotero, lo cual es innovador en el área de fitosanidad del estado de Guerrero.

Efecto de la temperatura en la sobrevida de salmonella en aguas grises de la provincia de Buenos Aires. Modificación de efecto con Escherichia coli.

Claudia Patricia Molinari (Universidad de Buenos Aires)

RESUMEN. *Salmonella sp y Escherichia coli son microorganismos patógenos e indicadores de contaminación fecal que pueden encontrarse en aguas provenientes de desechos domiciliarios conocidas como aguas grises. Las mismas son comúnmente originadas en duchas, piletas de cocina, etc. que se vierten separadas de las aguas negras. Cuando estas aguas se vierten en zanjas o canales a cielo abierto pueden ser fuente de transmisión de enfermedades debido a que los humanos tienen distintas formas de entrar en contacto directo con las mismas. En trabajos anteriores hemos estimado el riesgo de infección debido a Escherichia coli enteropatógena (Estimación del Riesgo Microbiológico con Escherichia Coli Enteropatógena con técnicas de SimulaciónC. Molinari et al,2012) y a Salmonella sp (Estimación del riesgo Microbiológico por Salmonella en aguas grises de la provincia de Buenos AiresC.Molinari et al, 2014) aplicando QMRA (Quantitative microbial risk assessment) a través de técnicas de simulación de Montecarlo. En este trabajo se evaluó el efecto de la temperatura en la sobrevida de Salmonella según la carga recogida en las muestras de la zona en estudio. La prueba se realizó en el laboratorio sometiendo a distintas temperaturas muestras de solución salina infectadas con el microorganismo, y realizando mediciones durante un mes. Por otra parte se realizó el mismo análisis pero agregando Escherichia Coli a las muestras infectadas a fin de evaluar el efecto combinado sobre la sobrevida de Salmonella.*

El análisis comparativo se realizó de dos maneras: a) ajustando una regresión para cada temperatura y comparando las pendientes . b)ajustando un modelo mixto. En ambos casos, ante la conclusión de rechazo de la igualdad de efectos, se realizaron comparaciones a posteriori. Se concluyó que la disminución de Salmonella se acelera al considerar 37°C respecto a temperaturas inferiores. A esa misma temperatura no pudo comprobarse un efecto acelerador generado por la infección con E. Coli.

ANÁLISIS DE SUPERVIVENCIA

Moderador: **Marco Scavino** (Universidad de la República)

Disability, quality of life and all-cause mortality in older Mexican adults: The effects of multimorbidity and frailty

Aaron Salinas Rodriguez (Instituto Nacional de Salud Pública)

RESUMEN. Background. Multimorbidity and frailty are relevant conditions among older adult population. Our aims were to assess the association of frailty and multimorbidity with the disability, quality of life and all-cause mortality as well as to analyze a potential effect modification between these conditions.

Methods. We performed a longitudinal analysis with older Mexican adults aged 50 years and older, using data collected from the WHO's Study on global AGEing and Adult Health Waves 1 and 2. Disability was measured using the World Health Organization Disability Assessment Schedule (WHODAS 2.0), and quality of life using the WHOQOL (WHO Quality of Life) instrument. All-cause mortality was determined by reviewing death certificates. Associations of frailty and multimorbidity with disability, quality of life and mortality were estimated using linear regression and Cox proportional hazards models.

Results. Multimorbidity assessed through three patterns (cardiopulmonary, vascular-metabolic, and mental-musculoskeletal) was associated with the three outcomes in this study. Cardiopulmonary and mental-musculoskeletal patterns increased the WHODAS mean score ($?=5.05$ and $?=5.10$, respectively) and decreased WHOQOL score ($?=-1.81$ and $?=-2.99$, respectively). Vascular-metabolic was associated with mortality ($HR=1.47$), disability ($?=3.27$) and quality of life ($?=-1.30$). Frailty was associated with mortality (pre-frail: $HR=1.48$ and frail: $HR=1.68$), disability (pre-frail: $?=5.02$, frail: $?=13.29$) and quality of life (pre-frail: $?=-2.23$, frail: $?=-4.38$). Interaction terms of frailty and multimorbidity were not statistically significant.

Conclusions. Multimorbidity and frailty are important predictors of poor health outcomes. These results highlight the importance of carrying out health promotion and prevention actions as well as specific interventions aimed at older adults who suffer from multimorbidity and frailty, in such a way that deleterious effects on health can be avoided.

Análisis de supervivencia para la estimación de la edad de inicio en el consumo de alcohol, marihuana y cocaína

Nora Beatriz Cadenas (Dirección Nacional de Observatorio Argentino de Drogas y Universidad Nacional de Tres de Febrero)

RESUMEN. Los datos del último estudio realizado en el año 2017 por La Dirección Nacional del Ob-

servatorio Argentino de Drogas (DNOAD) sobre “Consumo de Sustancias Psicoactivas en Población de 12 a 65 años”, permiten estimar que la tasa de consumidores recientes de alcohol, tuvo un aumento estadísticamente significativo en el año 2017 de aproximadamente 12 puntos con respecto al estudio, de idéntica metodología realizado en el año 2010, cuyos datos revelaron que el inicio precoz en el consumo de alcohol es un factor que aumenta la probabilidad de un consumo posterior de marihuana y/o cocaína, sobre todo entre los 25 y 34 años.

Objetivos:

Estimar, a partir de nuevos datos, la probabilidad de inicio del consumo de alcohol y evaluar dicho inicio como factor determinante en la transición al consumo posterior de marihuana y cocaína y comparar su variación entre los períodos de estimación 2010 y 2017. Material y Métodos

A partir de los datos de las Encuestas Nacionales sobre Consumo de Sustancias Psicoactivas en población de 12 a 65 años (2010-2017), se estimó la probabilidad acumulada de inicio en el consumo de alcohol y la transición, a partir de dicho consumo, al uso posterior de marihuana y/o cocaína. Para este análisis se utilizó la herramienta estadística de Análisis de supervivencia con el modelo multi-variado de Regresión de Cox. Se obtuvieron medidas de riesgo instantáneo de eventos en el tiempo (HR), con sus respectivos intervalos de confianza de 95

ISLP. Programa de alfabetización estadística internacional

Adriana Graciela D'Amelio Rey (Universidad Nacional de Cuyo)

RESUMEN. El ISLP nace en el 1994 desde el ISI (International Statisyical Institute) para estimular el desarrollo de habilidades cantitativas y su expansión en todo el mundo. A lo largo de los años se fueron incorporando acciones y actividades que se pretenden resumir en esta comunicación. Su misión es promover la alfabetización Estadística en todo el mundo y acompañar a través de los coordinadores de países en las actividades que ellos desarrollan. Grandes han sido los esfuerzos de mucha gente que con su trabajo desinteresado y comprometido ha colaborado y participado para que este programa se sostenga y expanda en el tiempo. Las fortalezas y debilidades de este programa como así su desarrollo a través del tiempo es lo que se quiere desarrollar en este trabajo.

El rol de la alfabetización

La alfabetización es un derecho humano fundamental y constituye la base del aprendizaje a lo largo de toda la vida. Por su capacidad de transformar la vida de las personas, la alfabetización resulta esencial para el desarrollo humano y social. El uso de la alfabetización para intercambiar conocimientos evoluciona constantemente, a medida que progresá la tecnología. Una comunidad alfabetizada es un colectivo dinámico, en el que se intercambian ideas y se suscitan debates. Una de las organizaciones que ha velado por más de 65 años por que la alfabetización conserve un carácter prioritario en los programas educativos nacionales e internacionales. , es la Unesco. En este sentido el programa del ISLP es fundamental en la interrelación con las sociedades de estadística, las universidades, las direcciones de estadística. Si un programa excluye los contextos sociales y culturales de una región o país este tiende a fracasar, es por ello que en esa diversidad el ISLP contempla la necesid

CONTRIBUCIONES LIBRES

CARTELES

Un modelo bayesiano para datos circulares basado en árboles de Polya

Daniel Eduardo Allard Oropeza (Universidad Autónoma Metropolitana)

RESUMEN. En los últimos años se ha vuelto a tener un desarrollo importante en las propuestas de metodologías estadísticas para analizar datos que representan direcciones, es decir, datos direcionales. Lo anterior, con énfasis en problemas que van desde el planteamiento de nuevos modelos estadísticos que describan el comportamiento probabilístico de estos datos, hasta metodologías para realizar inferencias en modelos de asociación como los modelos de regresión. Aunque la mayoría de los modelos para datos circulares (direcciones en dos dimensiones) son paramétricos, en muchas aplicaciones prácticas los datos observados poseen características que pueden no ser bien descritas a través de modelos paramétricos estándar.

Los datos direcionales, particularmente los datos circulares, pueden aparecer en modelos de regresión como la variable de respuesta. La teoría para modelos de regresión cuando la respuesta es una variable circular no se ha desarrollado completamente a pesar de que se pueden encontrar aplicaciones potenciales en varias áreas del conocimiento, particularmente en biología, geología y meteorología.

En este trabajo se presenta un modelo para describir datos circulares empleando particularmente procesos conocidos como Árboles de Polya. Se muestra la manera de hacer inferencias Bayesianas para todos los parámetros del modelo y se muestran las ideas para llevar a cabo inferencias en modelos de regresión donde la variable de respuesta es circular.

Estudio de las distribuciones a posteriori de los estimadores de regresión Lasso bayesiana y Lasso bayesiana robusta en grandes dimensiones

María Belén Allasia Bocca (Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario)

RESUMEN. Uno de los principales objetivos al analizar grandes volúmenes de datos es identificar aquellas variables que proveen información valiosa, de aquellas que no realizan aportes al objeto en estudio. Ante este contexto, la regresión LASSO (Tibshirani, 1996) se ha convertido en una alternativa ampliamente utilizada por su capacidad de contraer las magnitudes de los coeficientes, fijando en cero a aquellos cuya influencia no sea significante, realizando un procedimiento similar al de selección de variables. En el presente trabajo se realizó un estudio por simulación ajustando los modelos de regresión LASSO Bayesiana (Park & Casella, 2008) y una alternativa LASSO Bayesiana Robusta, a partir de distintos escenarios con diferentes cantidades de parámetros nulos y no nulos. En todos los casos, la cantidad de parámetros no nulos se varió entre un 10 % y 20 % del total de variables explicativas consideradas. Para cada simulación, se registraron los valores obtenidos como estimaciones de los parámetros del modelo y se construyeron las distribuciones empíricas a posteriori a partir de ellos. Se observó que ambos métodos de estimación son aproximadamente insesgados para la estimación de parámetros que se fijaron en valores nulos, y presentan un leve sesgo de para los no nulos. El estimador LASSO Bayesiano Robusto otorga resultados más consistentes, lo que garantiza

mejor precisión en las estimaciones que se realizan. Se considera que es la regresión LASSO Bayesiana Robusta es la opción más adecuada para el análisis de datos de esta índole, ya que además esta alternativa es menos sensible a la presencia de valores atípicos.

Hardy Weinberg equilibrium in Mendel's laws

David Almorza Gomar (Universidad de Cádiz)

RESUMEN. Mendel Matrix and Markov Process. Consider that the capital P represents the dominant factor, and lowercase p represents the recessive. From first and second Mendel's Laws, law of segregation and law of independent assortment respectively (Klug and Cummings, 2002), and using the order (PP, Pp, pp), it can be defined a transition matrix A. Transition matrix A will be called Mendel Matrix as it has been built from Mendel's Laws, and it will be used in a Markov process to proof the result presented in this paper. Hardy-Weinberg equilibrium set that in the absence of migration, mutation, natural selection and assortative mating, genotype frequencies at any locus are a simple function of allele frequencies (Wigginton, Cutler et al., 2005). Result: From any state, PP, Pp or pp, through Mendel Matrix, it is obtained a Hardy-Weinberg equilibrium situation. Proof: In a Markov process the distribution over states can be written as a stochastic row vector x (Brzezniak and Zastawniak, 2000). So, in the proof we use the Mendel Matrix and the Scientific Workplace 6.0 software. In this way we can conclude that from Mendel's laws we can obtain a situation of Hardy-Weinberg equilibrium, what was expected as the frequency of alleles does not change from generation to generation (in other words, the population does not evolve) and after one generation of random mating, offspring genotype frequencies can be predicted from the parent allele frequencies. References. Brzezniak, Z. and Zastawniak, T. (2000). Basic Stochastic Processes. Springer-Verlag (London). Klug, W.S. and Cummings, M. R. (2002). Essential of Genetics. Prentice Hall (New Jersey). Wigginton, J.E., Cutler, D.J. and Abecasis, G.R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. The American Journal of Human Genetics, vol. 76, issue 5, pp. 887-893.

Un modelo jerárquico temporal bayesiano para las máximas concentraciones de ozono de la Ciudad de México

Diana Alvarado Lima (Universidad Nacional Autónoma de México)

RESUMEN. Modelo Jerárquico Temporal Bayesiano para las Máximas Concentraciones de Ozono de la Ciudad de México

Autora: Diana Alvarado Lima, Facultad de Ciencias, UNAM. Directora de Tesis de Licenciatura: Dra. Lizbeth Naranjo Albarrán, Facultad de Ciencias, UNAM.

El presente trabajo de investigación pretende representar a la Distribución Generalizada de Valores Extremos (GEV) con una mezcla de escala de distribuciones uniformes y así ajustar a las concentraciones máximas de ozono de la Ciudad de México un modelo Bayesiano jerárquico temporal.

La representación de una distribución de una variable aleatoria en una mezcla de distribuciones permite obtener algoritmos Monte Carlo vía Cadenas de Markov más eficientes en la implementación de modelos estadísticos complejos. Existen diferentes representaciones de mezclas pero para el presente trabajo usaremos la clase de distribución de mezclas de escala de distribución uniforme, es similar a la representación de escala de distribución normal pero con la distribución normal reemplazada por una distribución uniforme cuyo soporte está determinado por el parámetro de mezcla.

Con esta representación de mezclas se ajustará la distribución GEV a las máximas concentraciones de ozono de la Ciudad de México, ya que hoy en día la contaminación del aire juega un papel importante en la salud de las personas y en particular, este contaminante provoca daños mayores en el sistema respiratorio. Como se busca explicar el comportamiento temporal de este contaminante durante los últimos 5 años, se presenta un modelo jerárquico temporal, donde el parámetro de localización tiene una estructura autoregresiva de orden $p=1$. Las inferencias correspondientes a este modelo se realizarán desde un enfoque Bayesiano ya que actualmente el uso de estos métodos nos permiten hacer análisis interesantes y más completos sobre los datos.

Bayesian quantile regression in stochastic frontier models

Angel Anibal Arroyo Hinostroza (Universidad Federal de Rio de Janeiro)

RESUMEN. We present a new class of models that use the power of Bayesian quantile regression to deal with stochastic frontier models. Compared with usual models that work with regression in the conditional mean, the proposal inherits the advantages of quantile regression, such as: robustness as it does not need to assume any distribution to the data, nor assume homoscedasticity, and also a better understanding since several quantiles provides more information about the response variable. In addition, the proposal allows to obtain statistically significant differences between technical efficiencies in lower quantiles.

In order to develop the inference we use the asymmetric Laplace distribution whose normal-exponential mixture representation allows to obtain known closed-form full conditional distributions. A Monte Carlo study is presented to empirically evaluate the methodology for Bayesian quantile regression. Two applications of the proposal are presented for cross-sectional and panel data sets. We compare the results obtained from our proposal with those obtained using conditional mean regression.

STATIS DUAL como alternativa de análisis de la interestructura del test de Inteligencia Emocional TMMS-24, en estudiantes de nivel superior.

Victoria Irandy Ballesteros Espinoza (Universidad de Colima)

RESUMEN. El estudio de la Inteligencia emocional en los estudiantes es un tema que se ha popularizado en los últimos años, esto se debe a que se ha encontrado evidencia de que una buena nota de calificaciones no es suficiente para que los alumnos tengan éxito en la vida, he aquí la importancia

de saber dominar nuestras emociones y conocer las de los demás para tener buenas relaciones interpersonales que son las que generan los vínculos sociales, es por ello que se realizaron tres estudios en tres años diferentes (2014, 2015 y 2016) en la Universidad de Colima obteniendo una muestra de 1045 alumnos de nivel superior, utilizando el test TMMS-24, que es una escala de 24 ítems que se divide en 3 dimensiones que son: Atención, Claridad y Reparación. Pero el objetivo principal de este estudio es saber si existe una estructura de covariación de los ítems del TMMS-24, derivado de que en los estudios publicados solo se limitan a la validación del instrumento, mediante técnicas factoriales, factoriales confirmatorios, análisis de consistencia interna, análisis de la varianza, pero ninguno contempla el análisis de las estructuras de covariación, por eso que se realizó un análisis con la ayuda del método STATIS DUAL, que es un método estadístico multivariante que permite representar la matriz de partida que corresponde al estudio en un tiempo; encontrando así que la estructura es comparable en los tres años calculando así una matriz compromiso que representa todas las matrices de partida de las fases del estudio, proyectado sobre ella los datos originales analizando así el comportamiento de los ítems y su evolución a lo largo del tiempo.

Métodos multivariantes HJ-biplot y cluster HJ-biplot como herramienta de inspección de los estilos de pensamiento en estudiantes de nivel superior.

Victoria Irandy Ballesteros Espinoza (Universidad de Colima)

RESUMEN. La evaluación de los estilos de pensamiento de los estudiantes se ha vuelto un tema de mucha importancia en los últimos años en las Universidades, esto se debe a que no todos los estudiantes aprenden de la misma manera, y para ellos se han realizado infinidad de pruebas psicométricas, en nuestro estudio una muestra de 349 alumnos de nivel superior de la universidad de colima de todas las áreas del conocimiento que contempla la Universidad, utilizando el test de Autogobierno mental (THINKING STYLES INVENTORY) de Robert J. Sternberg que consta de 104 ítems que conforman 13 subescalas agrupadas en 5 dimensiones para determinar los estilos predominantes de los estudiantes como se puede apreciar a continuación: Funciones: Legislativo, Ejecutivo y Judicial. Formas: Monárquico, Jerárquico, Oligárquico y Anárquico. Niveles: Global y Local. Alcances: Interno y Externo Inclinaciones: Liberal y Conservador Teniendo como base la teoría de autogobierno mental de R. Sternberg, respondiendo a la creciente necesidad de las instituciones educativas por encamarse hacia la calidad académica exige el reconocimiento de las características del desenvolvimiento del estudiante que permitan incorporar cambios en el proceso de enseñanza aprendizaje (Armas & Saavedra, 2015).

Nuestro objetivo de trabajo utilizar los métodos de análisis multivariante HJ-Biplot (Galindo, 1986) y Clúster HJ-Biplot (Vicente-Tavera, 1992) como herramientas de inspección de los Estilos de Pensamiento de los alumnos de primer ingreso de la Universidad de Colima. Es decir utilizaremos las técnicas multivariantes como instrumentos predictivos para analizar el comportamiento de los individuos y sus inclinaciones con relación a los estilos de pensamiento.

Vigilancia epidemiológica de las infecciones nosocomiales en una unidad de cuidados neonatales

Elia Barrera Rodríguez (Unidad de Innovación Clínica y Epidemiológica del Estado de Guerrero)

RESUMEN. De acuerdo como la Norma Oficial Mexicana, una IN es la reacción adversa a la presencia de un agente infeccioso o su toxina, que no estaba presente o estaba en periodo de incubación en el momento de ingreso del paciente al hospital. Las infecciones nosocomiales (IN) son un problema de importancia clínica y epidemiológica debido a que condicionan mayores tasas de morbilidad y mortalidad con un incremento en el costo social de años de vida potencialmente perdidos (Vigilancia Epidemiológica Hospitalaria (RHOVE), 2012). Como problema de salud pública, el Tercer Objetivo del Desarrollo Sostenible, propone poner fin a las muertes evitables de recién nacidos al 2030, en donde las IN son una de las principales causas de mortalidad neonatal, cuya tasa (41 %) eleva significativamente el índice general de mortalidad infantil (United Nations, 2010). En este trabajo se propone analizar la relación entre la variable respuesta binaria Infección Nosocomial Neonatal y las variables: Cuidados prenatales, condiciones al nacimiento del recién nacido y procesos invasivos posnatales. Se recolectó información de 1240 expedientes clínicos de neonatos atendidos en el periodo enero del 2012 a diciembre del 2014 en la Unidad de Cuidados Neonatales del Hospital de la Madre y el Niño Guerrerense.

Estudio de valores extremos con R: Una aplicación a temperaturas mínimas en Uruguay

Mathias Cardarello Fierro (Instituto de Estadística de la Facultad de Ciencias Económicas y de Administración - UdelaR)

RESUMEN. La modelización de los eventos extremos climáticos resulta de particular interés en la actualidad, debido al gran impacto que estos fenómenos producen en los sectores productivos. Estos últimos, que resultan especialmente sensibles a la variabilidad climática, adolecen en general de un conocimiento confiable sobre la ocurrencia de sucesos extremos y les es imprescindible asignarles ciertas probabilidades de ocurrencia. En consecuencia, existe una creciente demanda de investigación estadística seria sobre el hecho. En Uruguay, la producción agrícola y ganadera es una actividad económica clave ya que explica más del 70 % del valor total de sus exportaciones. La agricultura es un sector fuertemente afectado por las consecuencias de la variabilidad climática, entre las que se encuentran la ocurrencia de heladas agrometeorológicas que afectan la producción del sector. El objeto de interés es el estudio de las temperaturas mínimas en Uruguay. Son conocidos modelos espacio-temporales para establecer el nivel promedio de temperatura en una determinada región y un período de tiempo dado, pero muchas veces es necesario modelar las temperaturas extremas, esto es la cola de la distribución. Para la realización del trabajo utilizamos una base de datos compuesta por registros diarios de temperaturas mínimas, facilitada por el Instituto Uruguayo de Meteorología. Comparamos dos metodologías para la modelización de valores extremos. Se realizan predicciones de los niveles de retorno mediante el método del umbral y el método de valores extremos por blo-

ques. A nivel multivariado, la dependencia espacial a partir de datos provenientes de un conjunto de estaciones meteorológicas de Uruguay es analizada.

Variables latentes e inferencia bayesiana

Mario Enrique Carranza Barragán (Instituto Tecnológico Autónomo de México)

RESUMEN. Las variables latentes son variables aleatorias, en el sentido que presentan variabilidad, que en realidad nunca son observadas. El método de variables latentes consiste en introducir a un modelo, en el que es complicado hacer inferencia en los parámetros de interés, nuevas variables tales que, condicionado en ellas, se simplifique el modelo. Para ello debe establecerse la distribución de estas nuevas variables (latentes) y su relación con las variables observadas y los parámetros. Interesantemente, al fungir como puente entre los dos terminarán comportándose como parámetros y como variables. Afortunadamente, esto es de lo más natural en la inferencia Bayesiana.

La idea del método se ilustra con varios problemas de importancia en la práctica; incluyendo observaciones faltantes, mezclas de distribuciones, selección de variables en regresión, detección de outliers, modelación de variables ordinales y datos censurados.

Un modelo TRI de múltiples facetas para la evaluación del desempeño docente en el aula.

Karen Rosana Cordoba Perozo (Universidad Nacional de Colombia)

RESUMEN. Durante las últimas décadas las universidades se han preocupado por evaluar a sus docentes, un concepto que hace parte de esa evaluación es la percepción del estudiante, y bajo el supuesto de que los estudiantes aprenden más de los docentes altamente calificados se han usado evaluaciones sobre el rendimiento del desempeño docente en el aula. Se presentan cuestionarios con temas relacionados a la percepción que tienen los estudiantes de los docentes y de los cursos. Los cuestionarios relacionados pueden preguntar, por ejemplo, sobre características como el nivel de conocimiento del docente, la claridad en sus explicaciones, el cumplimiento de los cronogramas, percepción de dificultad del curso y organización del mismo. Cada ítem viene en una escala tipo Likert y los puntajes obtenidos se resumen generalmente con el cálculo de calificaciones promedio que involucran a todos los estudiantes en todos los ítems evaluados. Estas medidas promedio obtenidas se usan para comparar los desempeños de los docentes. Desde el punto de vista estadístico para una escala ordinal no tiene sentido hacer este tipo de comparaciones, pues estadísticamente no es lo mismo una diferencia de 1 a 2 en la escala a una diferencia de 3 a 4 aun cuando la magnitud en la diferencia sea de 1. Adicionalmente se han omitido detalles importantes al momento de la interpretación, como por ejemplo, el perfil del estudiante. Estos factores relacionados con el estudiante de cierta forma pueden influir en el concepto que tiene el mismo sobre el desempeño del docente y sobre la percepción global del curso. Según lo anterior se modelará el desempeño del profesor en el aula a partir de las evaluaciones hechas por sus estudiantes, mediante un modelo jerárquico de múltiples facetas (Linacre, (1989)). Las respectivas facetas serán: Los ítems relacionados con el desempeño

docente, el docente evaluado y los estudiantes evaluadores. Linacre, J. M. (1989). Many-facet Rasch measurement. Chicago: MESA Pres

Algunas metodologías de series de tiempo. Caso: Inbursa

Karla Cruz Montiel (Benemerita Universidad Autonoma de Puebla)

RESUMEN. En este trabajo se estudia la metodología de Box-Jenkins para realizar pronósticos a series financieras de precios de las acciones INBURSA, ya que nos permitira a corto plazo conocer como cambia el valor de la acción en algunos bancos y conocer indirectamente a que eventos o riesgos se pueden anticipar, la base con la que se cuenta es de 2013-2017 y se trabaja con diferentes metodologías.

Uso de redes bayesianas en el modelaje del comportamiento de un consumo sustentable

Carolina de Santiago Vázquez (Universidad Autónoma de Querétaro)

RESUMEN. El cambio climático ha hecho repensar los patrones de consumo en nuestras sociedades. Es importante comprender factores sociales que hagan que el comportamiento de consumo pueda modificarse hacia la sustentabilidad. Por ello el modelaje estadístico en este contexto es una herramienta útil. En este trabajo se estudia la teoría básica del modelaje mediante redes bayesianas y es aplicada a una encuesta realizada en Querétaro durante 2017. Se obtienen resultados y se comparan con resultados obtenidos a través de modelajes más tradicionales.

Estimación de los parámetros de un modelo lineal mixto bajo diferentes enfoques cuando los errores no cumplen los supuestos establecidos

Yadira Falcón Rosell (Universidad Autónoma de Guerrero)

RESUMEN. El modelo lineal mixto para datos longitudinales permite modelar la variabilidad presente en el comportamiento de un conjunto de unidades experimentales (sujetos) y en la unidad experimental a través del tiempo. de sujetos, e incluye En su formulación incluye parámetros fijos, comunes a toda la población, y parámetros aleatorios, los cuales permiten modelar la variabilidad presente en las unidades experimentales. específicos de cada unidad de muestreo. Este modelo se basa en supuestos de normalidad y varianza constante, tanto en el efecto como en el error aleatorio, sin embargo Además permite incorporar estructuras de correlación que son utilizadas para modelar la dependencia entre las observaciones a través de la matriz de varianza y covarianza de la distribución asumida para los errores. en su estructura de covarianza (coorelación) la correlación presente en las observaciones procedentes de la una misma unidad experimental. de muestreo presentan alta correlación originada por una serie de factores que los afecta, comunes a todas las mediciones procedentes

de una misma unidad, es por ello que diversos autores proponen diferentes estimaciones de este modelo, ya sea mediante el enfoque clásico o bayesiano. Dada que los errores ya no necesariamente se asumen independientes e idénticamente distribuidos, como generalmente se consideran, existen varios trabajos que proponen otras alternativas para la estimación de los parámetros involucrados en el modelo. En este trabajo se realiza un estudio por simulación para comparar el enfoque clásico y el bayesiano, utilizados para obtener las estimaciones de los parámetros del modelo. En el estudio de simulación se consideraron diferentes estructuras de correlación en los datos. con diferente nivel de correlación,. Finalmente se compararán los enfoques bayesiano y clásico del modelo mixto propuesto usando Para el enfoque bayesiano se consideran las prioris clásicas en que propone la literatura,, pcon lo que serán analizadas.

Cuantificación de incertidumbre en la teoría de control

Alexeis Figueroa Ferrer (Universidad Autónoma de Guerrero)

RESUMEN. En la actualidad, los sistemas de control son utilizados prácticamente en todas las disciplinas científicas y tecnológicas, desde sistemas muy simples como control de temperatura, hasta robots que controlan los procesos industriales en fábricas. Estos sistemas son estudiados por una rama interdisciplinaria de la ingeniería y las matemáticas, conocida como la teoría de control, que se ocupa de la modificación de sistemas dinámicos para obtener un comportamiento deseado en términos de un conjunto de especificaciones. En muchos fenómenos de la teoría de control, es común contar con una serie de mediciones o datos de algunas de las variables que forman parte de un modelo, el cual involucra ciertos parámetros desconocidos que representan características importantes del fenómeno bajo estudio. Si contamos con los datos, bajo incertidumbre, de esas variables, es posible plantear el problema inverso como un problema de inferencia estadística. En este trabajo identificaremos problemas concretos de la teoría de control y resolveremos el problema inverso como un problema de inferencia estadística desde el paradigma bayesiano.

Sobre la importancia de los val/vect propios en el análisis de datos, una implementación de Lanczos en R y aplicación a la segmentación de imgs en gran escala

Jose Antonio Garcia Ramirez (CIMAT)

RESUMEN. Si bien actualmente el término “big data” ha despertado el interés de la industria y la academia en el análisis de datos, independientemente de la definición de cada entidad (empresa o institución) del mismo término, existe una brecha entre el “big data” y datos en gran escala. En este trabajo se implementa el algoritmo de Lanczos, cuyas propiedades y aplicaciones son de provecho en estadística y las cuales se exponen brevemente en el trabajo, en paralelo y se aplica al problema de segmentar imágenes (utilizando el enfoque de Normalized cut) de dimensiones “grandes” (imágenes de tamaño mayores a 3000*3000) lo que implica manipular matrices de similaridad de dimensiones del orden de millones. La implementación, aunque es independiente del ambiente R puede ser empleada

en el mismo de manera que el usuario del entorno R pueda acceder a la implementación de manera sencilla. Esto es la base de un futuro trabajo con arquitecturas mixtas, usando arquitecturas populares como Hadoop y Apache Spark combinándolas con el poder de computo de los GPU's, esto nace con la finalidad de extender la actual biblioteca de algoritmos implementados en dichas arquitecturas populares al igual que hacer uso de ellas (de las cuales la documentación actual es poco basta al igual que el repertorio de algoritmos implementados en ellas en particular el análisis de imágenes está poco desarrollado en estas arquitecturas).

*Estudio de dispersión de un trazador en el lago de Zirahuén utilizando inferencia bayesiana:
Cadenas de Markov Monte Carlo*

Tzitlali Gasca Ortiz (Universidad Michoacana de San Nicolás de Hidalgo)

RESUMEN. El lago de Zirahuén es un lago de alta montaña, del tipo monomictico, oligotrófico y endorreíco que posee una gran variedad de flora y vegetación considerable, pero que se ve afectado por el deterioro de la explotación irracional de sus recursos, de las fuentes de polución de los asentamientos locales, el pastoreo de los animales y el cambio de uso de suelo, etc. La importancia de conocer el comportamiento de las variables que influyen en la dispersión de partículas es muy relevante, ya que el entendimiento de estas proporciona una herramienta para el manejo de los recursos y calidad del agua en el presente y futuro. La metodología utilizada en este estudio consiste en usar los resultados del modelo Delft3d que simulan la dispersión de un contaminante en el lago, además de observaciones de imágenes de alta resolución de un trazador. Mediante inferencia bayesiana con cadenas de Markov Monte Carlo (MCMC) se estima el parámetro de la viscosidad en respuesta a los datos observados. Se espera que los resultados de este estudio a largo plazo ayuden a establecer mejores prácticas y toma de decisiones para el manejo y conservación del Lago de Zirahuén.

*¿Cómo golden State Warrior llegó a ser campeón de la NBA? Uso de la estadística espacial
para analizar sus lanzamientos de campo*

Andrés Felipe Hernández Bustos (Universidad Nacional de Colombia)

RESUMEN. En la historia reciente los equipos de alto rendimiento de la mayoría de disciplinas deportivas se han empezado a interesar por análisis estadísticos que les permitan obtener mejores resultados. El baloncesto y la NBA como mayor referente no se han quedado atrás, y en los últimos años han construido potentes bases de datos que permiten a cada una de las franquicias de la liga realizar excelentes modelos estadísticos del comportamiento de sus jugadores y su equipo. Una rama de la estadística que poco se ha utilizado en dichos análisis, a pesar que puede ayudar a tomar decisiones importantes, es la estadística espacial. Por ello es de interés construir un procedimiento basado en herramientas dadas por la estadística espacial para llevar a cabo un análisis de los lanzamientos de campo en el baloncesto. Algunas de estas herramientas serán el de test de Kulldorff para

detectar clusters de alta incidencia emulando el proceso llevado a cabo por (Martinez 2013), así como herramientas proporcionadas por la teoría de patrones puntuales que será el fundamento teórico más fuerte usado en la investigación, tales como el cálculo de la intensidad en el campo de juego y la correlación especial presentada en el proceso. Por lo tanto , el objetivo principal será mostrar la aplicación de dicho procedimiento en el equipo Golden State Warriors y cada uno de sus jugadores durante las temporadas 2013-2014 y 2014-2015, teniendo en cuenta que en la primera temporada de estudio el equipo fue eliminado en la primera ronda de los play-offs mientras que en la segunda temporada con un nuevo coach y mostrando un patrón de lanzamiento particular logró el título de la NBA con una supremacía notable.

Esquemas de muestreo probabilístico para el conteo rápido bajo restricciones operativas.

Jeronimo Hernandez Mendoza (Universidad Autónoma Metropolitana)

RESUMEN. El Conteo Rápido (CR) es un procedimiento estadístico diseñado con la finalidad de estimar con oportunidad las tendencias de los resultados de una cierta elección. Se basa en la información de la votación de una muestra aleatoria de casillas electorales y las estimaciones se comunican a la población, en la noche del mismo día de la elección, en forma de intervalos de confianza.

Los encargados de organizar la elección son los Capacitadores y Asistentes Electorales (CAEs) quienes además de las múltiples actividades que deben realizar el día de la elección, también son responsables de recolectar la información de las casillas que hayan sido seleccionadas en muestra para el CR. Cada CAE tiene a su cargo, en promedio, 4 casillas. Esto plantea dos restricciones sobre el diseño muestral del CR:

1.- El día de la Elección la prioridad de los CAEs es ayudar a los funcionarios de casillas para que la elección fluya sin ningún contratiempo. 2.- Si un mismo CAE tiene más de dos casillas en muestra, este se verá presionado y no reportará la información de todas las casillas.

La solución del Comité Técnico del Conteo Rápido (COTECORA) para el Proceso Electoral 2018, en Elecciones Locales, fue:

a) Utilizar un diseño de muestreo estratificado, aquí se debe encontrar la estatificación que arroje el menor margen de error con el menor tamaño de muestra posible. b) Un tamaño de muestra con margen de error, en el peor de los casos, de 1 %. c) Del total de CAEs que apoyen al CR, mandando información de las casillas en muestra, mínimo el 80 % mandará información de sólo 1 casilla.

En este trabajo se desarrolla una propuesta que mejora la solución empleada por el COTECORA 2018. En otras palabras, utilizando un diseño de muestreo alternativo y con los mismos tamaños de muestra, usados por el COTECORA, se obtienen estimaciones con menor margen de error y se ejerce menor presión sobre los CAEs.

Uso de la Simulación en la solución de problemas de probabilidad: una experiencia de aprendizaje integradora

Roberto Hernández Ramírez (Universidad de Monterrey)

RESUMEN. En el proceso de solución de problemas de probabilidad los estudiantes tienen que confrontar las intuiciones en probabilidad producto de su experiencia de vida con el razonamiento probabilístico formal. Por desgracia la mayoría de las veces su intuición los engaña y los conduce a resultados erróneos o a dudar de la validez de los resultados correctos. El uso de la simulación para resolver problemas de probabilidad puede ser una buena estrategia didáctica que además de promover el aprendizaje de los elementos básicos de probabilidad, el razonamiento probabilístico y desarrollar las nociones de modelado de procesos aleatorios, también permite que los alumnos tengan la oportunidad de integrar otros conocimientos como los principios de programación y lograr una experiencia de aprendizaje integradora. El objetivo de esta presentación es compartir las experiencias que se han tenido al usar esta aproximación en un curso de probabilidad y estadística básico. Se presentan y describen algunos problemas asignados de complejidad graduada que involucran el uso de herramientas como Raptor, Geogebra y Excel. Se discuten también algunas dificultades en la implementación de esta estrategia didáctica.

Introducción a swirl en la enseñanza de programación y las ciencias de datos en R

Nery Sofía Huerta Pacheco (Universidad Veracruzana)

RESUMEN. Tema (clasificación principal): Cómputo estadístico

El presente trabajo se muestra como parte introductoria del paquete swirl en R project que es una herramienta para elaborar guías de estudio y pruebas relacionadas con la enseñanza en temas de programación básica, así como tópicos centrados en las ciencias de datos. Para ello, se expondrá de forma concisa los elementos necesarios para crear contenidos didácticos que apoyen a los estudiantes de estadística desde una forma interactiva en el proceso de aprendizaje. Además de proporcionar ejemplos prácticos de asignación, tipos de datos, expresiones aritméticas y lógicas, estructuras ?secuenciales, estructuras selectivas, estructuras repetitivas, programación modular, arreglos unidimensionales y bidimensionales, evaluación y presentación de resultados aplicados a nivel licenciatura.

Estimación de Razones de Momios para Cáncer Cervicouterino utilizando modelos de regresión binaria aplicado a una clínica de la ciudad de Durango.

Edgar Felipe Lares Bayona (Instituto de Investigación Científica de la UJED)

RESUMEN. En epidemiología el concepto de riesgo es utilizado para evaluar la posibilidad de que la enfermedad se presente más comúnmente con una condición específica. El factor de exposición es la condición que tienen los individuos con mayor riesgo de enfermar. El Cáncer Cervicouterino

(CaCu) se estudia en una población de la clínica de atención del Instituto de Investigación Científica de la ciudad de Durango. Mundialmente el CaCu ocupa los primeros lugares en morbilidad y mortalidad en países en desarrollo. La estimación de razones de momios con la regresión logística es la herramienta estadística más utilizada en el área de la medicina, sin embargo, existen más modelos de regresión para datos binarios que pueden ser aplicados para la estimación del riesgo. Una de las principales dificultades de la población estudiada fue detectar una gran cantidad de pacientes con diagnóstico negativo para CaCu contra una pequeña cantidad de pacientes positivos, en específico, se observó una prevalencia del 0.6 % para CaCu. El ajuste de un modelo de regresión logística resulta en estimaciones sesgadas, lo que conlleva a buscar modelos apropiados que permitan obtener mejores estimaciones. Para esto, se estimaron razones de momios calculados con diferentes modelos de regresión binaria con funciones de enlace simétricas y asimétricas, un modelo de regresión logística para eventos raros y un modelo de regresión binaria Bayesiano semiparamétrico. De acuerdo a los resultados se encontraron tres factores de riesgo acorde al contexto para CaCu siendo el virus del papiloma humano estadísticamente significativo. En un análisis posterior mediante la recodificación de las variables regresoras ordinales a variables categóricas dicotómicas se encontró qué si bien ésta estrategia ayuda a determinar una mayor cantidad de riesgos acordes al contexto para CaCu, encontrar variables con riesgo significativo no acordes al contexto son considerados como sesgos no propios del análisis estadístico.

Dispersión de clorofila-a en la costa de Jalisco, enfoque probabilístico.

Andrea Patricia Manrique Cantillo (Universidad de Guadalajara)

RESUMEN. Dadas las escasas investigaciones que se tienen en la Costa de Jalisco relacionadas con la dispersión de Clorofila-a, la conectividad que se presenta en el sistema marino costero y la interacción de dichas partículas con las variables climatológicas que se presentan en el área de estudio, es necesario plantear metodologías que permitan identificar y definir cuál es el comportamiento de estas partículas de clorofila. Se entiende la importancia ecológica y económica que tiene la clorofila, puesto que, en un futuro, el conocer el movimiento en la zona costera puede ser de utilidad en actividades como la pesca por la relación que presenta respecto a la productividad primaria. En este estudio se presenta un enfoque Bayesiano dados los resultados de un modelo numérico que simulan el movimiento de partículas por varios forzamientos para inferir la estimación de los parámetros que dan lugar a lo observado durante el año de estudio de las imágenes satelitales que se tienen de la NASA.

Teoría de respuesta al ítem para la medición de la salud oral en un estudio uruguayo

Franklin Fernando Massa Mandagar (Universidad de la República)

RESUMEN. En los estudios epidemiológicos es práctica habitual trabajar con variables binarias que reflejan la presencia de determinadas enfermedades, las que a su vez pueden estar asociadas con

otro conjunto de variables, que en general se asumen como factores de riesgo de las primeras. En el ámbito de los estudios epidemiológicos referidos a la salud oral, es común indagar sobre la relación entre la presencia de algunas patologías y ciertas características de los participantes del estudio a través de modelos lineales generalizados (generalmente regresión logística). No obstante ese tipo de análisis suele ser llevado a cabo de manera separada para cada variable de interés y en ningún momento se obtiene una medida que resuma el estado de cada participante. En este estudio se propone el uso de modelos de IRT (específicamente el modelo de Rasch) ya que estos permiten el análisis conjunto de varias variables obteniendo como subproducto una valoración individual, que en este caso se interpreta as "salud oral". Adicionalmente, el análisis presentado aquí extiende el modelo de Rasch incluyendo un predictor lineal que permite indagar sobre algunos los factores determinantes de la tendencia de los individuos a padecer las distintas patologías. Los resultados de este estudio confirman el efecto de la edad sobre la disposición a presentar enfermedades orales.

Estrategias de enseñanza que se utilizan cuando se aplican secuencias didácticas con tic en bioestadística

Emma Graciela Montañez (Facultad de Ciencias Exactas y Naturales- Universidad Nacional de Catamarca)

RESUMEN. La enseñanza de Bioestadística se basa en el concepto de sentido estadístico cuyas componentes son: comprensión del estudiante, competencia de análisis de datos, razonamiento estadístico. Este último, es el más difícil de transmitir y lograr en los alumnos en un semestre, y ello se evidencia en el bajo rendimiento obtenido en los últimos años. Esta preocupación motiva a analizar las estrategias de enseñanza que se utilizan cuando se trabaja con secuencias didácticas con Tecnologías de la Información y la Comunicación en la cátedra Bioestadística del Profesorado y Licenciatura en Biología, basadas en competencias. El enfoque de investigación considerado fue mixto: cuantitativo (descriptivo-comparativo) y cualitativo (análisis de contenido). Los resultados del análisis descriptivo, con esta nueva metodología, superaron a los de años anteriores. Las secuencias didácticas basadas en competencias abarcaron, entre otras estrategias, el uso del Dropbox con diversos materiales on-line y la aplicación de Tecnología. Del análisis de contenido de la aplicación de las secuencias didácticas se deduce que ellas permitieron reflexionar e incorporar nuevas estrategias de enseñanza y aprendizaje, evidenciando que se puede aprovechar mejor el tiempo y distribuirlo de forma tal que permita desarrollar todo el programa de la asignatura en profundidad. Si bien se detectaron algunas dificultades en los estudiantes al aplicar ciertas estrategias de estudio, el uso de las secuencias didácticas con Tecnología –como recurso didáctico- ha potenciado el proceso de enseñanza y aprendizaje, favoreciendo la construcción del conocimiento en los alumnos. Palabras claves: Estrategias, Enseñanza, Secuencias didácticas, TIC, Bioestadística.

*Análisis estadístico para la diferenciación microscópica de tres variedades de arvejas
comercializadas en la ciudad de Humahuaca*

Judith Montenegro Brusotti (Universidad de Buenos Aires. Facultad de Farmacia y Bioquímica)

RESUMEN. En la ciudad de Humahuaca, ubicada en el noroeste de la República Argentina, se comercializan, en la actualidad, y en abundancia, tres variedades de arvejas. Este tipo de legumbres suele presentarse en el mercado local como semillas enteras, trozadas o en harina y es por lo que, la diferenciación, es necesaria en los distintos niveles. Macroscópicamente son bien diferentes, pero a nivel microscópico no lo son. El objetivo de este trabajo es poder determinar, a través del análisis estadístico de los datos obtenidos del material analizado micrométricamente, la variedad de la que proviene la harina o los pequeños trozos que se comercializan. Para ello, se comparan las longitudes y anchos de las esclereidas que componen el tegumento de cada una de las variedades de semillas. La metodología empleada para la preparación del material en estudio consiste en hidratar las arvejas, retirar los tegumentos y realizar con ellos un disociado fuerte para poder separar las células. Se observa bajo microscopio óptico y, con un ocular graduado, se mide el tamaño de las esclereidas. El resultado del análisis mostró 2 tipos de esclereidas (esclereidas columnares y osteoesclereidas) que se encontraban en las tres variedades de arvejas, algunas con un amplio rango de tamaño que hace necesario un tratamiento estadístico para validar diferencias. Al realizar el análisis estadístico no se evidencia variabilidad entre la esclereida columnar de dos de las variedades, pero se evidencian diferencias significativas con el otro tipo celular (osteoesclereida). Se concluye a partir de este resultado que ante una presentación comercial de harina o pequeños trozos de arveja es posible determinar la variedad de la que proviene.

Modelos dinámicos lineales: una aplicación para el pronóstico del nivel de ozono

Nayeli Montiel Rodríguez (Universidad Nacional Autónoma de México)

RESUMEN. En los últimos años el crecimiento demográfico y junto con éste el desarrollo industrial que ha venido experimentando la Ciudad de México, han ocasionado que problemas como la contaminación atmosférica y el deterioro de la calidad del aire se vean agravados. Por lo que en la actualidad el estudio de medidas para contrarrestar sus efectos ha tomado mayor relevancia.

El ozono ha formado parte de diversos estudios e investigaciones. El ozono es uno de los contaminantes ambientales más dañinos, por los efectos adversos con los que se relaciona: el ausentismo escolar y problemas de salud pública, que van desde síntomas respiratorios hasta el incremento de tasas de morbilidad y mortalidad por enfermedades relacionadas con los sistemas respiratorio y cardiovascular.

En este trabajo se pretende realizar un análisis de series de tiempo de observaciones univariadas bajo un enfoque bayesiano. Se analizarán las concentraciones de ozono registradas en la red automática de monitoreo ambiental de la Ciudad de México, para la estación Pedregal (PED), la cual hemos identificado como la estación con mayores concentraciones de este contaminante en el año 2017. Y

como marco de referencia se utilizarán los modelos dinámicos lineales (MDL), por su versatilidad para el ajuste de modelos variantes en el tiempo (West and Harrison, 1997). Siendo nuestro objetivo establecer un modelo de regresión lineal, tomando como covariable la temperatura, para el pronóstico temporal del nivel de ozono.

Referencias:

Petris, G.; Petrone, S.; Campagnoli, P. (2009) Dynamic Linear Models with R., New York, USA. Springer.
West, M.; Harrison, J. (2004) Bayesian Forecasting and Dynamic Models, New York, USA. Springer Series in Statistics, segunda edición. Springer-Verlag.
Sahu, S. K.; Bakar, K. S. (2011), “A comparison of Bayesian models for daily ozone concentration levels”. ELSEVIER, Statistical Methodology, 9, 144-157.

Thin-plate spline partially varying-coefficient model under elliptical distributions

Magaly Sofia Moraga Cárdenas (Universidad Austral de Chile)

RESUMEN. In this work we extend varying-coefficient models with normal errors to elliptical contoured errors in order to allow distributions with heavier and lighter tails than the normal ones. This class of models includes all elliptical contoured continuous distributions, such as Student-t, Pearson VII, power exponential and logistic, among others. Estimation is performed by maximum penalized likelihood method and by using smoothing splines. Thin plate splines are implemented for modeling the spatial trend. A real data set previously analyzed by using varying-coefficient models with normal errors is reanalyzed under varying-coefficient models with heavy-tailed errors.

Evaluación de la colaboración y el aprendizaje de diagramas de dispersión en estudiantes con tecnología

Maria del Pilar Morales Ramirez (Universidad Veracruzana)

RESUMEN. El aprendizaje colaborativo parte de un modelo de enseñanza en las aulas desde hace prácticamente una década, esto es que los estudiantes trabajen de forma colectiva para participar o compartir información, en la cual discuten ideas y construyen su propio conocimiento en base a la teoría de Vygotsky en el 2000. Actualmente, existen diversas formas de mediar la colaboración en las aulas y más cuando se hace uso de recursos tecnológicos, ya que comúnmente se requiere que los estudiantes trabajen con sus compañeros dado los escasos recursos materiales de cada escuela. El objetivo de esta investigación radica en identificar la existencia de índices colaborativos que influyen en el proceso de aprendizaje de estudiantes de secundaria definidos de manera inicial, en grupos por los estilos de aprendizaje. Para ello, se realizó un estudio exploratorio con estudiantes de nivel básico de la asignatura de matemáticas en la Escuela Secundaria General No. 2 “Julio Zárate” de la Ciudad de Xalapa, Veracruz. Las unidades de estudio fueron estudiantes de primer grado a los cuales se evaluó el nivel de conocimiento del tema sobre diagramas de dispersión a través de

pruebas estandarizadas iniciales y finales durante la interacción con un Sistema Tutor Inteligente llamado “Scooter”. Por otro lado, durante el proceso de adquisición de conocimientos, se utilizaron dos instrumentos estandarizados validados para definir los estilos de aprendizaje y la colaboración. Los resultados encontrados se evaluaron por medio de técnicas no paramétricas para la comparación de muestras pareadas, así como medidas de asociación, mostrando que existen índices de influencia colaborativa y estilos predominantes que afectan al aprendizaje.

Representaciones sociales sobre la Estadística como disciplina científica: un primer acercamiento

José Juan Muñoz León (Universidad Veracruzana)

RESUMEN. En este trabajo se reportan elementos del conocimiento de sentido común sobre la Estadística como disciplina científica, que se mantienen, pierden y ganan entre estudiantes de último año de bachillerato, estudiantes de primer semestre y de semestres avanzados de la Licenciatura en Estadística de la Universidad Veracruzana. Tomando como referencia la Teoría de las Representaciones Sociales (Moscovici, 1976), en este estudio se analiza el producto en el que se concreta una representación social (Moscovici, 1976): conocimiento socialmente elaborado y compartido (conocimiento de sentido común), que tiene una finalidad práctica y que concurre a la construcción de una realidad común a un conjunto social (Jodelet, 1989). Con el objetivo de contribuir al diseño de pautas para posicionar socialmente a la Estadística en diferentes contextos escolares se ha aplicado un instrumento con estudiantes de primer semestre en la Licenciatura en Estadística, en donde se ha observado que la palabra más frecuente asociada a *Estadística.es Matemáticas*, seguida de números, gráficas, datos y administración. Es interesante que la población de estudiantes de semestres avanzados de la misma licenciatura incluye palabras como Análisis, Datos, Gráficas, Matemáticas, entre otras, más no incluye Cómputo, Diseño, Interpretación, Resultados, Desiciones, y otras palabras que aparecen al definir a la Estadística desde el conocimiento teórico. En el instrumento, además de palabras asociadas, se ha solicitado a los estudiantes que formen equipos y construyan una definición sobre la Estadística que sea aceptada al menos por los miembros del grupo. Se ha transcrita la construcción y el análisis del discurso favorece la misma tendencia: para estudiantes de Estadística la Estadística es Matemáticas. ¿Es favorable esta tendencia?

Visualización y análisis mediante series temporales de la siniestralidad vial del Uruguay en el período 2013-2017

Matías Muñoz Wolf (Universidad de la República)

RESUMEN. En el período 2013 – 2017 hubo 326.464 siniestros de tránsito, en los cuales resultaron 124.723 heridos leves, 19.134 heridos graves y 2527 fallecidos. Debido a la magnitud que alcanza esta problemática en Uruguay, se ha llevado a cabo durante los últimos años la implementación de nuevas leyes, medidas y campañas de concientización con el fin de mejorar la seguridad vial. La UNASEV

(Unidad Nacional de Seguridad Vial) publica regularmente datos nacionales sobre siniestralidad. En base a estos datos es que se presenta el análisis de aquellos siniestros en los que los afectados resultaron lesionados sin llegar a constatar el fallecimiento dado el gran impacto social y económico que esto genera al país. El presente trabajo tiene como objetivo la descripción de la siniestralidad vial en el Uruguay para el período 2013-2017 utilizando el análisis de la serie temporal. Se busca generar herramientas de visualización interactiva mediante los paquetes shiny y dygraph del software estadístico R, que permitan estudiar el impacto de las diversas campañas de prevención realizadas en el período de estudio así como generar un modelo de predicción para los años venideros.

*Investigación del hábito tabáquico y noción del daño en pacientes oncológicos y allegados.
Repercusión de las leyes de protección*

Myriam Carmen Nuñez (Universidad de Buenos Aires)

RESUMEN. El tabaquismo constituye la principal causa de muerte evitable en los países desarrollados. Los cinco principales riesgos para la salud están relacionados con el comportamiento y la alimentación: índice de masa corporal alto, bajo consumo de frutas y hortalizas, inactividad física, tabaquismo e ingesta excesiva de alcohol. Estos factores causan el 30 % de las muertes por cáncer. El consumo de tabaco es el más importante: causa 22 % de las muertes mundiales por cáncer en general y 71 % de las muertes por cáncer de pulmón. Los objetivos de este trabajo son: conocer la prevalencia y características del tabaquismo, y la conciencia de daño en pacientes con cáncer (GP) y sus allegados (GF). 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. Conocer la repercusión que han tenido las nuevas leyes de protección contra el tabaco en la conciencia y prevalencia del hábito. Ambas encuestas se realizaron en el Hospital Italiano de Buenos Aires (HIBA). Se obtuvieron muestras de 364 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. Si bien se hallaron diferencias entre los períodos estudiados que hablan del beneficio que produjeron las nuevas leyes de protección, consideramos que es fundamental seguir trabajando para que estas diferencias sean más notorias.

Análisis de Series de Tiempo Irregulares

Cesar Andres Ojeda Echeverri (Pontificia Universidad Católica de Chile)

RESUMEN. En este trabajo se propone una nueva clase de procesos estocásticos llamados procesos de medias móviles autorregresivos irregularmente espaciados de orden uno que permiten el tratamiento

y el análisis de series de tiempo irregulares. Se establece su definición formal y sus características para posteriormente estudiar a partir de métodos de monte carlo el comportamiento de los estimadores por máxima verosimilitud de sus parámetros. Se presenta ilustraciones con datos reales y simulados.

A Bayesian nonparametric hypothesis testing procedure for paired samples

Luz Adriana Pereira Hoyos (Pontificia Universidad Católica de Chile)

RESUMEN. We propose a Bayesian hypothesis testing procedure for comparing the distributions of paired samples. The procedure is based on a flexible model for the joint distribution of both samples. The flexibility is given by a mixture of Dirichlet processes. Our proposal uses a spike-slab prior specification for the base measure of the Dirichlet process and a particular parameterization for the kernel of the mixture in order to facilitate the posterior inference. The joint model allows us to derive the marginal distributions and testing whether they differ or not. Our procedure exploits the correlation between both samples, relaxes the parametric assumptions and allows us to detect the possible differences across the entire distribution. We provide illustrations with real and simulated data sets. Additionally, a Monte Carlo simulation study is provided, such study compares the performance of our proposal with some traditional alternatives.

Análisis del proceso Código Rojo usando modelos de ecuaciones estructurales

Eduardo Perez Castro (Universidad Autónoma de Guerrero)

RESUMEN. A pesar de los logros notables en la reducción de mortalidad materna y neonatal mundial, cada día mueren alrededor de 830 mujeres en el mundo por complicaciones relacionadas con el embarazo o el parto. Se ha reportado que tres cuartas partes de estas, podrían ser prevenibles si las mujeres tuvieran acceso a servicios obstétricos de alta calidad. El Código Rojo (CR) es un procedimiento clínico dirigido a gestantes con situaciones de emergencia obstétrica que incluye la identificación de la morbilidad, manejo inicial y de las condiciones mórbidas para su remisión. El embarazo induce un gran número de cambios fisiológicos, sin embargo, esta misma condición, dificulta la atención de la paciente y pueden dar lugar a daños secundarios al embarazo; por ejemplo, el daño hematológico, renal, hepático, pulmonar y neurológico. Los datos se obtuvieron del Hospital de la Madre y el Niño Guerrerense que da atención a madres y niños con riesgo alto. Las mujeres embarazadas se presentaron al servicio de urgencias del HMNG con cuadro clínico compatible con emergencia obstétrica de pacientes atendidas durante los meses de septiembre de 2013 a diciembre de 2015. En el periodo hubo 224 expedientes clínicos útiles. En este trabajo se propone usar modelos de ecuaciones estructurales para analizar la relación entre la variable latente respuesta Estado de salud de la madre y del recién nacido y las variables latentes independientes son: primer estado hemodinámico de la madre, segundo estado hemodinámico, tratamientos, antecedentes gineco obstétricos y gasto económico.

Redes Bayesianas para la predicción de ozono en la Ciudad de México

Carlos Samuel Pérez Pérez (ITAM)

RESUMEN. La calidad del aire en la Ciudad de México es un problema importante que afecta la salud de millones. La contaminación del aire se encuentra entre las más altas del mundo, con un promedio diario significativo de contaminantes primarios como hidrocarburos, óxidos de nitrógeno y monóxido de carbono. Estos provienen principalmente de las emisiones del transporte y la industria que, una vez expuestas al sol, sufren reacciones químicas que producen una variedad de contaminantes secundarios como el ozono. Los que están en mayor riesgo al respirar aire contaminado incluyen personas con asma, niños, adultos mayores y personas activas, especialmente trabajadores, al aire libre.

Existe una necesidad real de modelos que puedan pronosticar los niveles de contaminación con anticipación de varias horas, o incluso días, para tomar medidas de emergencia tales como son los planes de contingencia. Las áreas donde no existen mediciones también requieren estimaciones para tomar medidas preventivas y así reducir la exposición a este peligro para la salud. Este trabajo presenta resultados que permiten comprender el problema de la calidad del aire en la Ciudad de México al determinar los principales factores que contribuyen a la contaminación a través de datos horarios. Solo se toma en cuenta información estadísticamente relevante gracias a los algoritmos de aprendizaje para obtener la estructura inicial de los fenómenos y, como subproducto, se plantean interacciones principales de contaminantes y otros factores atmosféricos. Las redes bayesianas se usaron para modelar interacciones entre variables y revelar la importancia de los factores en la estimación de los niveles de contaminantes. La aplicación de redes bayesianas es útil para obtener una mejor comprensión de los fenómenos en los cuáles solo es posible una observación parcial y brindar una noción de precisión accionable para la política ambiental.

Proyección de cantidad de llamadas de Contact Center

Luisa Jacqueline Reyes Medina (Universidad del Bío-Bío)

RESUMEN. En este trabajo se ocuparán datos reales proporcionado por el Contact Center de una empresa del área de telecomunicaciones de la zona, la cual está en pleno desarrollo, lo que otorga una dificultad a la predicción a la que se pretende llegar. El problema es la alta variabilidad en la cantidad de llamadas del Contact Center según las contingencias, días feriados, días de suspensión de clientes, etc. donde el principal objetivo es estimar la cantidad de llamadas para el mes siguiente, y así poder asignar los turnos a los trabajadores. Para resolver nuestro problema se diseñará pronósticos de demanda mediante el método de Suavizamiento Exponencial, la metodología de Box-Jenkins (ARIMA) y el método Espacio Estado, para luego comparar sus resultados y decidir que procedimiento ofrece el mejor resultado de predicción.

Modelos ocultos de Markov para series de tiempo financieras: una aplicación en índices bursátiles

Yadira Rivas Godoy (Facultad de Ciencias, UNAM)

RESUMEN. Un modelo oculto de Markov (Hidden Markov Model, HMM), es un modelo estadístico en el cual el sistema a estimar es un proceso de Markov con parámetros desconocidos, en donde la ocurrencia de los estados está asociada con una distribución de probabilidad y las transiciones entre los estados están gobernadas por un conjunto de probabilidades de transición. Un HMM es un caso particular de las redes bayesianas dinámicas. Estos modelos se utilizan para reconocer datos secuenciales tales como datos financieros. Los datos de series de tiempo financieras son una secuencia en un periodo de tiempo determinado, los cuales pueden ser precios o rendimientos de activos financieros, donde la volatilidad es una característica inherente de las mismas. Por lo tanto, es necesario usar otros modelos distintos a los modelos de series de tiempo tradicionales que suponen varianza homocedástica. Usualmente, para este tipo de datos se propone usar un modelo GARCH, que es una generalización del modelo ARCH, e incluso un EGARCH, en los cuales se toman en cuenta las características de la volatilidad, así como su comportamiento asimétrico frente a las alzas y bajas de los precios de un activo financiero. Sin embargo, estos modelos no siempre resultan en el mejor ajuste. Por consiguiente, se estudiarán los procesos de HMM por la utilidad que tendrán en este tipo de datos, específicamente en la parte de predicción de series de tiempo financieras, donde se podría pensar que los datos se generan por algún proceso estocástico subyacente desconocido para el público. Para finalizar, haremos una aplicación de lo anterior, estimando HMM para indicadores bursátiles, que son los más volátiles y desafiantes del mercado de valores. Teniendo en cuenta que, un indicador bursátil es el índice de referencia que se conforma de un conjunto de valores cotizados en una bolsa de valores (e.g. el Índice de Precios y Cotizaciones, IPC, en nuestro país). Elliott RJ, Aggoun L, Moore JB (1995) Hidden Markov Models.

Estimación del coeficiente de Gini utilizando distribuciones tipo fase

Luz Judith Rodriguez Esparza (Universidad Autónoma Chapingo)

RESUMEN. En este trabajo, se considera una nueva metodología para calcular el Coeficiente de Gini utilizando distribuciones tipo fase, en particular, se utilizan las distribuciones de momento tipo fase. Para estimar dicho coeficiente, se calculan los estimadores de máxima verosimilitud de la distribución tipo fase, por lo que se utiliza el algoritmo EM, algoritmo que se ha utilizado ampliamente en cuestiones de estimación de esta clase de distribuciones. Ilustramos la eficiencia del método propuesto calculando el coeficiente de Gini y la curva de Lorenz para México considerando los años 1995, 2000 y 2005.

Medidas de dependencia y dependencia extrema no positiva utilizando cópulas; aplicación a datos financieros

Daniel Román Álvarez (Facultad de Ciencias)

RESUMEN. Se hace un estudio sobre los principales resultados de cópulas y su conexión con las medidas de dependencia extrema y funciones de dependencia extrema, tanto positiva como no positiva. Se aplican los resultados a tres series financieras para generar en mejor modelo D-via trivariado con su respectivo análisis de bondad de ajuste.

Estudio de las notificaciones por VIH/Sida para el periodo 2013 - 2017 en el servicio de salud Talcahuano a través del uso de modelos GAS

Javiera Salazar Baez (Universidad del Bío-Bío)

RESUMEN. En Chile, al año 1984 se registra el primer caso de VIH/SIDA, desde aquella fecha ha existido un creciente interés por entender el comportamiento y detener el avance de esta enfermedad actualmente declarada como epidemia. En los últimos años, considerando la explosiva alza de nuevos casos de infección por VIH, se ha fijado la atención en este problema de salud pública, que registra al año 2016 un aumento de un 34 % en el total de casos nuevos de VIH en adultos respecto a lo observado el año 2010, para el año 2016 se tiene registro de aproximadamente 5.000 casos nuevos por este virus, convirtiendo a Chile en el país latinoamericano que presentó la mayor alza en esta materia.

Para la realización de este estudio se cuenta con información de las Enfermedades de Notificación Obligatoria (ENO), las que se basan en el reporte, tanto del sector público como privado. Se utiliza específicamente el total de notificaciones mensuales de las enfermedades con código CIE10 (B20-B24): Enfermedades causadas por el Virus de la inmunodeficiencia humana (VIH) y (Z21): Estado de infección asintomática por el virus de la inmunodeficiencia humana [VIH], a contar del mes de enero del 2013 a diciembre del año 2017 de las personas con residencia en las comunas de; Talcahuano, Hualpén, Tomé y Penco, esto con el fin de modelar y proyectar la media de este proceso Poisson a través del uso de una clase de modelos de serie temporales llamados GAS (Generalized Autoregressive Score) los cuales se encuentran basados en el score de la verosimilitud.

Consumo de medios de comunicación de universitarios veracruzanos

Marisol Sánchez Cano (Universidad Veracruzana)

RESUMEN. En la actualidad los medios de comunicación siguen siendo parte fundamental para mantener a sus audiencias con un mejor acceso a la información, sin embargo, la manera de interactuar con estos se ha visto transformada drásticamente en los últimos años, es decir, la relación audiencia-medios ha sido afectada por diferentes situaciones sociales y/o avances tecnológicos, por lo que es interesante realizar investigaciones en este sentido que permitan conocer como se está dando este

fenómeno en poblaciones específicas. Debido a lo anterior, se realizó una encuesta de consumo de medios a estudiantes y docentes universitarios, para conocer las relaciones existentes entre la comunidad universitaria y los medios, llevándose a cabo en todo el estado de Veracruz, el cual, fue dividido en zonas para realizar un muestreo estratificado, el levantamiento de dicha encuesta se realizó durante los meses de enero y febrero del presente año. Para fines de la realización del cartel que se presentara, se seleccionaron solo algunas variables de interés de dicha encuesta, de inicio se mostraran algunos resultados de un análisis descriptivo que permita visualizar cuales son los medios de comunicación que más utilizan tanto de carácter académico como de carácter privado y la credibilidad que le dan a la información que en esos medios proporcionan, y posteriormente los resultados obtenidos de un análisis multivariado que dé como resultado un agrupamiento entre variables que permita dar una interpretación condensando el máximo posible de información.

Una herramienta interactiva en R para la enseñanza del análisis de conglomerados

Enrique Santibáñez Cortés (Universidad Autónoma Chapingo)

RESUMEN. El análisis de conglomerados consiste en la formación de grupos de individuos similares a partir de mediciones de algunas de sus características, donde la noción de similitud se formaliza en con el uso de una métrica. Recientemente, ha esta técnica ha cobrado importancia por su uso en el análisis de datos con el enfoque de ciencia de datos y de aprendizaje de máquina. El objetivo de la presentación es mostrar una aplicación web interactiva, desarrollada con el paquete Shiny del lenguaje de programación estadística R, que puede servir como auxiliar en la enseñanza del análisis de conglomerados. La aplicación permite comparar a través de ejemplos los resultados que se obtienen al aplicar distintos algoritmos para la formación de conglomerados y calibrados con distintos parámetros.

*Modelación espacial de la plaga Sigakota (*Mycosphaerella fijiensis*) en cultivos de plátano (*Musa spp*) del estado de Guerrero*

Juan Elías Solís Alonso (Universidad Autónoma de Guerrero)

RESUMEN. Los bananos y plátanos (*Musa spp.*) se cultivan en más de 120 países alrededor del mundo, es una importante fuente de carbohidratos en la dieta de millones de personas, sobre todo en países tropicales, además representan el cuarto cultivo de mayor importancia, después del maíz (*Zea mays*), trigo (*Triticum aestivum*) y arroz (*Oryza sativa*), a nivel mundial. La variedad de bananos y plátanos que se cultivan en México es amplia, entre ellas destacan el Plátano Tabasco o Roatán, Enano Gigante, Criollo, Valery, Dominicano, Manzano y el plátano Macho, Morado y Pera. Guerrero ocupa el octavo lugar con una producción anual de 50 a 60 toneladas por hectárea. La Sigakota (*Mycosphaerella fijiensis*), una de las principales enfermedades del cultivo, provoca una considerable reducción en el área foliar de la planta y en consecuencia los racimos y los frutos tienen un menor peso en comparación con las plantas sanas. Este problema finalmente se ve reflejado en la producción total.

El objetivo de este trabajo fue modelar la distribución espacial de la enfermedad en el municipio de Tecpán de Galeana, ubicado en Costa Grande Guerrero, a través de un proceso espacial gaussiano estacionario. Para la modelación de la estructura de dependencia espacial de los datos se utilizó la función del semivariograma; y para la interpolación del proceso el método de kriging ordinario y universal. Esta investigación permitió identificar los puntos de mayor gravedad de la enfermedad en la zona de estudio, lo cual es muy importante a la hora de implementar las medidas de control para la Sigakota negra. La base de datos, que consta de 1827 puntos de muestreos georreferenciados, comprende los monitoreos de julio a diciembre del año 2017, seis meses en total. La información fue proporcionada por SAGARPA. Palabras clave: Interpolación espacial, Semivariograma, Proceso espacial gaussiano.

Comparación de 2 métodos estadísticos para identificar grupos en una encuesta de satisfacción estudiantil

Elena Vernazza Mañan (Instituto de Estadística, Universidad de la República)

RESUMEN. En este trabajo se estudian las principales características de la construcción de la satisfacción estudiantil, en los cursos de grado de la Facultad de Ciencias Económicas y de Administración, Universidad de la República, Uruguay, a través de la utilización y comparación de dos técnicas de análisis de datos multivariantes: análisis de clusters (en particular, cluster jerárquico) y análisis de clases latentes. Los datos utilizados para la aplicación presentada en este trabajo provienen de una encuesta aplicada sobre una muestra de estudiantes de grado de la Facultad, en el año 2009. Dicho cuestionario, presenta una estructura de bloques de preguntas; el primero contiene las variables que permiten realizar un caracterización sociodemográfica de los estudiantes. Por otra parte, se presentan las variables del modelo ECSI (European Customer Satisfaction Index) que serán las utilizadas para la caracterización de la satisfacción estudiantil. Las 6 variables observadas/manifiestas que se consideran como insumos para la construcción y caracterización de la satisfacción estudiantil son: expectativas (E) de los estudiantes al ingresar al centro de estudios, la imagen (I) que tienen de éste, la calidad de la enseñanza recibida (CSA) y de los servicios brindados (CSF), las necesidades y deseos personales con respecto a la Facultad (ND) y el valor percibido (VP). Los resultados presentados surgen, por un lado, de considerar que efectivamente existe una variable no observada/latente que refiere a la satisfacción estudiantil y que ésta queda definida, a partir de la interacción de las 6 variables manifiestas, por 4 clases latentes. Por otra parte, en lo que refiere a los clusters, se propone agrupar a los estudiantes en 3 grupos, a partir del análisis de los resultados que surgen de clusterizar a partir de la metodología Ward. Se presenta, además, la comparación de los resultados obtenidos con ambas técnicas.