

## SESIONES PLENARIAS

### Sesión plenaria 1.

**Alejandro Jara. Department of Statistics. Pontificia Universidad Católica de Chile**

#### **R package for Bayesian Non- and Semi-parametric Analysis (DPpackage 1.1-6)**

Abstract:

This package contains functions to perform inference via simulation from the posterior distributions for Bayesian nonparametric and semiparametric models. Although the name of the package was motivated by the Dirichlet Process prior, the package considers and will consider other priors on functional spaces. So far, DPpackage includes models considering Dirichlet Processes, Dependent Dirichlet Processes, Dependent Poisson- Dirichlet Processes, Hierarchical Dirichlet Processes, Polya Trees, Linear Dependent Tailfree Processes, Mixtures of Triangular distributions, Random Bernstein polynomials priors and Dependent Bernstein Polynomials. The package also includes models considering Penalized B-Splines. Currently the package includes semiparametric models for marginal and conditional density estimation, ROC curve analysis, interval censored data, binary regression models, generalized linear mixed models, IRT type models, and generalized additive models. The package also contains functions to compute Pseudo-Bayes factors for model comparison, and to elicitate the precision parameter of the Dirichlet Process. To maximize computational efficiency, the actual sampling for each model is done in compiled FORTRAN. The functions return objects which can be subsequently analyzed with functions provided in the coda package.

### Sesión Plenaria 2.

**Massimiliano Caporin. Department of Statistical Sciences. University of Padova. Italy**

#### **Do jumps matter in Realized Volatility modeling and forecasting? Empirical evidence and a new model**

Abstract:

Building on an extensive empirical analysis I investigate the relevance of jumps and signed variations in predicting Realized Volatility. I show that properly accounting for intra-day volatility patterns and staleness sensibly reduces the identi\_ed jumps. Realized Variance decompositions based on intra-day return size and sign improve the in-sample \_t of the models commonly adopted in empirical studies. I also introduce a novel speci\_cation based on a more informative decomposition of Realized Volatility, which o\_er improvements over standard models. From a forecasting perspective, the empirical evidence I report shows that most models, irrespective of their exibility, are statistically equivalent in many cases. This result is confirmed with different samples, liquidity levels, forecast horizons and possible transformations of the dependent and explanatory variables.

### **Sesión Plenaria 3.**

**Yongcheol Shin. Department of Economics, University of York. United Kingdom**

#### **Canonical Correlation-based Model Selection for the Multilevel Factors**

##### **Abstract**

A great deal of research effort has been devoted to the analysis of the multilevel factor model. To date, however, limited progress has been made on the development of coherent inference for identifying the number of the global factors. We propose a novel approach based on the canonical correlation analysis to identify the number of the global factors. We develop the canonical correlations difference (CCD), which is constructed by the difference between the cross-block averages of the adjacent canonical correlations between factors. We prove that CCD is a consistent selection criterion. Via Monte Carlo simulations, we show that CCD always selects the number of global factors correctly even in small samples. Further, CCD outperforms the existing approaches in the presence of serially correlated and weakly cross-sectionally correlated idiosyncratic errors as well as the correlated local factors. Finally, we demonstrate the utility of our framework with an application to the multilevel asset pricing model for the stock return data of 12 industries in the U.S.

### **Sesión Plenaria 4.**

**Esther Ruiz Ortega. Department of Statistics. Universidad Carlos III de Madrid**

#### **Factor extraction using Kalman filter and smoothing: This is not just another survey**

##### **Abstract**

Dynamic Factor Models, which assume the existence of a small number of unobserved latent factors that capture the comovements in a system of variables, are the main "big data" tool used by empirical macroeconomists during the last 30 years. One important tool to extract the factors is based on Kalman filter and smoothing procedures that can cope with missing data, mixed frequency data, time-varying parameters, non-linearities, non-stationarity and many other characteristics often observed in real systems of economic variables. This paper surveys the literature on latent common factors extracted using Kalman filter and smoothing procedures in the context of Dynamic Factor Models. Signal extraction and parameter estimation issues are separately analyzed. Identification issues are also tackled in both stationary and non-stationary models. Finally, empirical applications are surveyed in both cases.

## **Sesión Plenaria 5.**

**David B. Dunson. Statistical Science, Trinity College of Arts & Sciences, Duke University**

### **Avoiding over-interpretation in learning structure in data: Uncertainty quantification in clustering**

Machine learning tools are routinely used for finding structure in data, a problem referred to as unsupervised learning - in contrast to supervised learning problems that typically focus on prediction. Canonical examples of unsupervised learning include k-means clustering and PCA. In general, the vast majority of unsupervised learning methods used in practice lack any characterization of uncertainty. For example, in clustering of data using k-means, we obtain a single clustering without any ability to assess whether other clusterings are also consistent with the available data. This can naturally lead to over-interpretation and a propagation of errors in the literature. This talk focuses on approaches to avoid this problem through the use of Bayes and generalized Bayes methods. We start with a case study to single cell genomics data in which clusters are meant to correspond to different cell types. A typical approach that is used in this setting is to apply PCA in a first stage to reduce dimension and then apply k-means. As a first alternative, we consider a model-based clustering approach using mixture models. We show that such methods (which are very popular in the Bayesian literature) face a pitfall in high dimensions that was previously unknown, and we propose a solution to this pitfall. As a second alternative, we propose a generalized Bayes approach that allows one to directly characterize uncertainty in loss-based unsupervised learning methods, such as k-means.

## CONTRIBUCIONES LIBRES

### **Ponencia P1: Daniel Ventosa-Santaularia, CIDE & Banco de México**

A simple and powerful skewness test

We propose a distribution symmetry test based on the statistical equality of two variance estimators: the first variance estimate uses the r.v. realizations that are inferior to the sample mean whilst the second one uses those greater than the sample mean. We prove that, under (i) an i.i.d. assumption, and (ii) a symmetric distribution, both estimates are equal asymptotically and thus a simple equality test can be build. The test statistic is normally distributed under the null hypothesis of symmetry whilst it diverges at a rate  $T^{1/2}$  (where  $T$  is the sample size) under the alternative. We show, using Monte Carlo finite-sample simulations that the proposed is powerful and bears a good level-power relationship. Finally, a brief empirical application is presented. Applied statisticians are familiar with mean (centrality or location) and variance (dispersion or spread), but less so with higher moments, such as skewness (and kurtosis) that refer to the shape of the distribution. Skewness is a salient property of a distribution as it may offer a particularly relevant insight in fields such as finance, where it may reveal, for instance, investors' preference for positive yields

### **Ponencia P2: Carlos Guerrero de Lizardi, Facultad de Economía, UNAM**

Granger revisited: t values and the empirical OLS bias with stationary and non- stationary time series using Monte Carlo simulations

The estimation method most frequently used in applied econometrics is that of ordinary least squares. On one hand its theoretical characteristics are "ideal"; on the other, Clive Granger, using as a background the nonsense correlations introduced by Yule (1926) and a framework based on balanced and unbalanced equations, explored the fact that, in the presence of stationary and non-stationary time series, standard ordinary least squares inference is misleading. We graphically address the empirical distribution of the bias of the ordinary least squares estimator and the inconvenience of the use of standard errors which underestimate its true variation. Also, we show evidence that Granger's path-breaking ideas have not been sufficiently taken up by the specialized literature –which, besides, we highlight using a "measurement in economics" perspective. Last but not least, taking into account the golden rule of sciences, we point out that our quantitative exercises are replicable to the extent we fully share our codes, and we use the open-access database of the seminal paper written by Nelson and Plosser (1982).

### **Ponencia P3. Edilberto Nájera Rangel, Universidad Juárez Autónoma de Tabasco**

Comparación de algunos métodos para estimar por intervalo los parámetros de la distribución gamma

En este plática se comparan los métodos siguientes para estimar por intervalo los parámetros de la distribución gamma: método a través de la función de verosimilitud perfil, método de los intervalos tipo Wald, método fiducial y métodos a través de la distribución final bayesiana cuando la distribución a priori es: la no informativa de Jeffreys, la de referencia cuando el parámetro alfa es de interés y el parámetro beta es de ruido, y la de referencia cuando beta es de interés y alfa de ruido. La comparación se hace a través de las probabilidades de cobertura y las longitudes esperadas de los intervalos respectivos, las cuales son estimadas usando simulación de Montecarlo. Palabras clave: Inferencia bayesiana, inferencia fiducial, intervalos de confianza, longitud esperada, probabilidad de cobertura, verosimilitud perfil.

#### **Ponencia P4. Arrigo Coen Coria, University of Wisconsin-Madison**

Identificación de redes filogenéticas bajo un modelo de pseudoverosimilitud

Los avances actuales en la obtención de las secuencias genéticas traen consigo nuevas preguntas para la bioinformática. El análisis de estas secuencias permite entender los patrones evolutivos de la vida y con ello avanzar en distintos campos de las ciencias biológicas. Por ejemplo, el conocer la evolución de un virus contribuye a la obtención de vacunas para combatirlo. Ya que existen organismos cuya evolución proviene de procesos de hibridación, es necesario estudiar modelos de redes filogenéticas. Al respecto, en esta plática se hablará de las condiciones de identificación de un modelo de redes filogenéticas basado en el modelo de coalescencia. Este modelo es ajustado utilizando pseudoverosimilitud, lo cual es motivado por la complejidad de la verosimilitud de una red con respecto al número de especies y al número de eventos de hibridación que contiene.

#### **Ponencia P5. Jesús Arroyo. Johns Hopkins University**

Detección de comunidades en muestras de redes

Resumen de la ponencia. Indic: La detección de comunidades en una red es un problema de aprendizaje no supervisado que consiste en agrupar vértices con patrones de conexiones similares. Aunque el estudio de métodos para encontrar comunidades en una red ha sido un área de investigación activa, en muchas aplicaciones es común encontrar muestras de redes con posibles covariables adicionales. En neurociencia, por ejemplo, la estructura de comunidades en la red de conexiones cerebrales usualmente corresponde a regiones responsables de diferentes funciones cerebrales. La conectividad cerebral es diferente en cada persona, y datos de muestras de redes cerebrales individuales pueden ayudar a entender la relación entre la conectividad y variables de interés, como características individuales o desórdenes cerebrales. En esta plática presentaremos métodos espectrales para encontrar comunidades en muestras de redes, y mostraremos como la estructura de comunidad puede ser usada como regularización para predecir variables y encontrar comunidades supervisadas de acuerdo a las variables de interés.

#### **Ponencia P6. Miguel David Álvarez Hernández. INFOTEC**

Un análisis de la complejidad del sistema geo-electoral mexicano

El objetivo principal de la ponencia es presentar una metodología de análisis que permite cuantificar la complejidad del sistema geo-electoral mexicano, asociada a la realización de las tareas de Integración, Identificación, Mantenimiento y Actualización del Registro Federal de Electores. Para lograr este objetivo, se elabora un conjunto de Índices de Complejidad Electoral, a escala municipal y seccional, sustentados en información procedente de diversos indicadores geoespaciales, demográficos y estructurales. Los resultados estadísticos obtenidos a partir de los índices se complementan con la elaboración de una Tipología de Complejidad Electoral, la cual permite clasificar a los municipios y secciones electorales en función de sus grados de complejidad electoral.

### **Ponencia P7. Alberto Padilla.**

Una fórmula para la varianza del muestreo por conglomerados en términos de la correlación intraclase y la varianza poblacional entre elementos

En el muestreo por conglomerados el concepto de homogeneidad juega un papel fundamental en diversos aspectos de dicho diseño. Dicha homogeneidad se mide con el coeficiente de correlación intraclase para el caso del muestreo por conglomerados en dos etapas con tamaños iguales y puede expresarse en términos de las variaciones entre y dentro de conglomerados. En este diseño muestral, la fórmula para la varianza del estimador del promedio, usando muestreo aleatorio simple en ambas etapas, se expresa en términos de variaciones entre y dentro de conglomerados. Esto hace que sea difícil evaluar directamente el efecto de la correlación intraclase en la varianza del estimador. En esta plática se exhibe una fórmula para la varianza del estimador en el muestreo por conglomerados en términos de la varianza poblacional entre elementos y el coeficiente de correlación intraclase. Se muestran varios resultados del efecto de la homogeneidad en la varianza del estimador, así como una expresión exacta para la correlación intraclase en términos del efecto del diseño.

### **Ponencia P8. Gladys Linares. BUAP**

Teoría de respuesta al ÍTEM: Análisis de un cuestionario de conciencia ambiental

El objetivo de este trabajo es realizar el análisis de un cuestionario de Conciencia Ambiental a través de la Teoría de Respuesta al Ítem, lo que permitirá contar con una herramienta para la medición y distribución en categorías de la Conciencia Ambiental en los ámbitos preuniversitarios y universitario, descrita por las diferentes dimensiones que definen este concepto: cognitiva, afectiva, conativa y activa. La investigación se desarrolló durante el curso 2019, a una población de 60 alumnos de nivel medio superior del Tele bachillerato Manuel Velasco Suarez, en Pijijiapan, Chiapas. Después de la aplicación del cuestionario, se caracterizaron los reactivos o ítems (preguntas) del mismo a través de los criterios propuestos por la TRI, verificándose a través del examen de la correlación entre el ítem y el puntaje total del instrumento y de las Curvas Características del ÍTEM de las cuatro dimensiones del cuestionario, que es posible medir la Conciencia Ambiental de estudiantes de este nivel.

### **Ponencia P9. Eder Rubelo.**

Desarrollo de perfiles analíticos.

La importancia de desarrollar perfiles que genere conocimiento con los datos de las PyME

### **Ponencia P10. Jesús Francisco Escalante Euán. Universidad Autónoma de Yucatán**

Enfoques metodológicos para valorar el riesgo

La gestión del riesgo (risk management), es un enfoque estructurado que incorpora la incertidumbre relativa a una amenaza, vinculada una secuencia de actividades inherentemente humanas que incluyen la evaluación del riesgo, así como otras estrategias para su mitigación. El objetivo, es reducir diferentes riesgos relativos a un ámbito preseleccionado. Su clasificación es diversa, por ejemplo: las amenazas por factores asociados al medio ambiente, la tecnología, los errores humanos, las organizaciones, inclusive la política. La ponencia sentará las bases del marco lógico

implementado, y que en definitiva pueda orientar a los tomadores de decisión hacia la adopción del enfoque de gestión del riesgo en sus proyectos, ya que proporcionan una señal potencialmente importante en cuanto a dónde invertir los recursos ante contextos clave. Palabras clave: riesgo, análisis multivariante, simulación.

#### **Ponencia P11. Ricardo Hoyos. Banco de México**

Índice de Liquidez en el mercado de Bonos Mexicanos y el efecto de la crisis por Covid

Se presenta una construcción simple de un índice de liquidez para el mercado de bonos en México, y una posible interpretación del mismo con base en árboles binarios

#### **Ponencia P12. Jorge Alberto Nájera Salmerón. UAM Azcapotzalco**

La importancia de definir diferentes estadísticos para medir el Riesgo de un Portafolio de Inversión.

Se presentaran 4 diferentes criterios (la desviación estándar, el VaR, el VaR Condicional y el parámetro de un modelo ARCH-GARCH) que son utilizados en la Elección de Portafolios de Inversión para definir el Riesgo de una Inversión; especificando sus principales características, y concluyendo en que la mejor estrategia de inversión se toma al considerar cada uno de estos modelos.

#### **Ponencia P13. Mauricio Villanueva. ITAM**

Análisis de la burbuja del precio de Bitcoin bajo un modelo log-periódico con término de error AR-GARCH.

Como trabajo de tesis de licenciatura trabajé con un modelo de series de tiempo, relativamente poco conocido, para describir el crecimiento de una burbuja en el valor de un activo financiero y estimar la fecha más probable del estallido de la misma: el modelo log-periódico. En este caso, para modelar el precio de la criptomoneda Bitcoin, analizando su comportamiento a lo largo del año 2017.

#### **Ponencia P14. (Ponencia invitada) Osvaldo Espin García. Princess Margaret Cancer Centre & University of Toronto.**

Uso de marcadores poli génicos en estudios de dos fases para secuenciación masiva dirigida.

En las últimas décadas, los estudios de asociación genómica (EsAG) han identificado cientos de miles de asociaciones entre polimorfismos de un solo nucleótido (PsSN) y una multitud de fenotipos. Sin embargo, estas asociaciones únicamente localizan regiones genómicas de interés. Idealmente, los investigadores buscar secuenciar masivamente todas las variantes genéticas en estas regiones para todos los participantes. No obstante, los costos de secuenciación pueden ser elevados, sobretodo cuando el tamaño de muestra es grande. Considerando lo anterior, los estudios de dos fases proveen de un marco estadístico el cual permite reducir costos derivados de secuenciar regiones genómicas de interés. Esta reducción de costos se logra mediante la identificación y

selección de un subconjunto de participantes (S2) que provee mayor información para realizar la inferencia y para los cuales se realiza la secuenciación masiva. Es importante notar que la selección de S2 se realiza utilizando los datos colectados en la fase 1, es decir, fenotipo (Y) y PsSN (Z). El objetivo de la fase 2 es hacer inferencia entre las variantes secuenciadas (G) y el fenotipo (Y). Por lo tanto, un estudio de dos fases corresponde a un problema de datos incompletos, donde el mecanismo de la información faltante es especificado por el diseño. Por otro lado el uso de marcadores poli génicos (MPGs) se ha extendido ampliamente en años recientes. Los MPGs son sumas ponderadas de PsSN que buscan cuantificar la susceptibilidad genotípica que tiene un individuo de manifestar un determinado fenotipo y donde los ponderadores son típicamente parámetros de regresiones realizados en EsAG que incluyen cientos de miles de participantes. En este proyecto, evaluamos el uso de MPGs para informar la selección de S2 en estudios de dos fases mediante simulaciones Monte Carlo y finalizamos con una ilustración usando un estudio de cohorte de nacimiento finlandés, The North Finland Birth Cohort of 1966.

### **Ponencia P15. Osvaldo Díaz Álvarez. Instituto Politécnico Nacional**

Efecto de terapias modificadoras de la enfermedad sobre la calidad de vida de pacientes con esclerosis múltiple: Revisión sistemática y meta-análisis en red

El objetivo del estudio fue comparar el efecto sobre la calidad de vida relacionada con salud (HRQOL) de distintas terapias modificadoras de la enfermedad (TME) para el tratamiento de pacientes con esclerosis múltiple remitente recurrente (EMRR), por medio de un meta-análisis en red (NMA). Se realizó una revisión sistemática incluyendo ensayos clínicos aleatorizados que reportaran resultados sobre HRQoL en pacientes con EMRR. El NMA se desarrolló en R, a través de la paquetería “netmeta”. Se localizaron 16,123 publicaciones, analizando 6 estudios en 9 NMA para los siguientes cuestionarios: EQ-5D determinado por VAS y por índice de salud; MSIS-29 en dominios físico y psicológico; SF-36 en su dominio físico. Se observaron diferencias significativas entre alemtuzumab vs IFN  $\beta$ 1a en EQ-5D VAS. Así como entre IFN  $\beta$ 1a y daclizumab en MSIS-29, tanto en rubro físico como psicológico. Los demás NMA no mostraron diferencias estadísticamente significativas entre los TME evaluados.

### **Ponencia P16. Edilberta Tino Salgado. Universidad Autónoma de Guerrero**

Ensayo clínico aleatorizado para el control de la presión arterial en pacientes puerperales con preeclampsia grave: Uso de la ANOVA de medidas repetidas

Se llevó a cabo un ensayo clínico longitudinal, aleatorizado, simple ciego, de 42 puérperas con preeclampsia grave. Los pacientes fueron aleatorizados en dos grupos: el grupo experimental (n = 21) recibió diltiazem (60 mg) y el grupo de control (n = 21) recibió nifedipina (10 mg). Ambos fármacos se administraron por vía oral cada 8 horas. Se registraron y analizaron las presiones arterial sistólica, diastólica y media, así como la frecuencia cardíaca al inicio y después de 6, 12, 18, 24, 30, 36, 42 y 48 horas. (ANOVA de dos factores de medidas repetidas). Las medidas de resultado primarias fueron todos los parámetros de presión arterial mencionados anteriormente. Las medidas de resultado secundarias incluyeron el número de episodios de hipertensión e hipotensión junto con la duración de la estancia en la unidad de cuidados intensivos.

### **Ponencia P17. Fernando Marmolejo Ramos. University of South Australia**

A new distribution to model reaction time data

Data can take different distributional shapes but some shapes are more common in certain research fields. Reaction times (RTs) from participants are typically collected in experimental psychology and neuroscience studies. The distribution of RTs tends to follow a positive skew and the ex-Gaussian (EG) distribution is the usual distribution used to model such data. However, it is not unlikely RTs can take normal or even negatively skewed shapes. In those cases, the EG distribution doesn't provide a good fit. We propose a distribution able to accommodate positive, normal and even negative shapes, hence becoming a flexible distribution to model RT data.

### **Ponencia P18. María de los Dolores Sánchez Castañeda. ITAM**

Modelo jerárquico espacio-temporal de homicidios en México

Se pretende conocer el comportamiento del fenómeno de homicidios en México analizando las relaciones espacio-temporales durante el periodo 2000-2012 a nivel estatal, considerando las características demográficas, sociales y culturales específicas de cada entidad. El análisis se realizará mediante modelos lineales espaciales generalizados bajo el enfoque Bayesiano. Se determinarán las entidades con mayor incidencia de homicidios y los factores de riesgo considerando la perspectiva de género. Estos modelos serán de utilidad para la toma de decisiones y focalización de políticas de prevención, así como para el seguimiento de los esfuerzos en seguridad pública y en los efectos de las intervenciones en cada estado.

### **Ponencia P19. Luz Judith Rodríguez Esparza. Universidad Autónoma de Aguascalientes**

Índice de Riesgo al Suicidio en México utilizando Twitter

Identificar de manera temprana indicios de riesgo de cometer suicidio por personas depresivas, de forma que se les pueda proporcionar la atención especializada pertinente. Diversos estudios se han centrado en estudiar las expresiones en redes sociales, donde los usuarios vierten sus emociones, para determinar si muestran indicios de depresión o no. Sin embargo, han dejado de lado la cuantificación del riesgo de cometer suicidio. Por ello, este artículo propone un nuevo índice para identificar el riesgo al suicidio en México. La propuesta de este índice se construye a través de la minería de opinión utilizando Twitter y el Proceso Jerárquico Analítico.

### **Ponencia P20. Mónica Aranzazú Medina. Unidad Académica de Matemáticas**

Índice Sobre Percepción de Seguridad

El objetivo es desarrollar un sistema de índices de percepción sobre la seguridad pública para los municipios de Guadalupe, Zacatecas y Fresnillo (Estado de Zacatecas, México) mediante una encuesta confiable aplicada a muestras representativas durante al menos dos periodos en el año. La hipótesis que se plantea es la siguiente: los índices de percepción de seguridad en los municipios de Zacatecas y Guadalupe son similares pero distintos a los del municipio de Fresnillo porque los dos primeros están fusionados de tal forma que conforman una zona metropolitana, dado que comparten los mismos problemas de inseguridad. Los conceptos utilizados es la Estadística Descriptiva, tablas de contingencia, ponderaciones, índices y finalmente mapas perceptuales.

## **Ponencia P21. Nadia Magally Lázaro Jiménez. Universidad de las Américas Puebla**

### Optimización de código en R

En esta plática se mostrarán las ventajas y desventajas de optimizar código de R utilizando vectorización o Rcpp. Se expondrán tanto ejemplos simples como complejos que sean de fácil entendimiento pero que también muestren el amplio alcance de optimización que se puede llegar a conseguir con una buena reestructuración de código. Asimismo, se señalarán las limitantes que tienen estos métodos y se discutirá la razón de su nulo funcionamiento. Se presentarán los errores más comunes al utilizar Rcpp, así como las paqueterías y librerías dentro de R que son útiles para la resolución de éstos.

El objetivo de esta ponencia es proporcionar la información mínima necesaria para que el espectador pueda plasmar, en sus propios códigos, los puntos antes mencionados, además de que pueda identificar con claridad en qué casos se puede realizar esta optimización y en qué casos se requiere otra metodología.

## **Ponencia P22. Sergio Mota. IIMAS, UNAM**

### Diferentes enfoques para la detección de anomalías.

La detección de anomalías ha sido utilizada desde hace tiempo para la localización y posterior eliminación de observaciones anómalas en un conjunto de datos. En las últimas décadas ha aumentado la investigación en detección de anomalías, ya que estas son una fuente de información invaluable respecto al tema de interés en una investigación. Las anomalías nos brindan datos respecto a fallos estructurales, averías mecánicas, comportamiento fraudulento, error humano o simplemente mostrar desviaciones naturales en observaciones obtenidas en una investigación. En esta exposición se mostrarán diferentes enfoques para la detección de anomalías, estos enfoques van desde la estadística hasta algoritmos de aprendizaje automático.

## **Ponencia P23. Manuel Durazo. ITAM**

### Análisis semántico para revelar tratamientos contra el COVID19

En la ponencia se resumirá el trabajo publicado en <https://www.kaggle.com/mdurazob/cord-19-vaccines-networks-and-bibliometrics> donde el uso de redes y parámetros de centralidad nos permitió identificar tendencias en una cantidad masiva de publicaciones científicas sobre el tema. Estas tendencias coinciden hasta hoy con el camino tomado por la ciencia para desarrollar la vacuna contra el COVID19. El análisis de redes y la inteligencia artificial permite descubrir conocimiento ahí donde no había sido generado.

## **Ponencia P24. Diana Alvarado Lima. Banco de México**

### Análisis de puntos de cambio en volumen de forwards sobre divisas peso/dólar

Este trabajo presenta una aplicación para el análisis de puntos de cambio en un conjunto de datos. El problema general se refiere a la inferencia de un cambio en la distribución de un conjunto de variables ordenadas en el tiempo. El enfoque implica la estimación paramétrica y no paramétrica,

tanto del número de puntos de cambio como de las posiciones en las que ocurren. El enfoque es general, y en el caso paramétrico asume que la distribución de los datos se distribuye normal mientras que, en el no paramétrico, no hace supuestos sobre la naturaleza de la distribución de los datos. El procedimiento de estimación se basa en la agrupación jerárquica y la aplicación de algoritmos tanto de división como de aglomeración. El método se utiliza para explorar el momento y el número de puntos de cambio en el volumen de importe concertado en operaciones de forwards sobre divisas peso/dólar tanto de compra como de venta, que se concertaron en el mercado de derivados mexicano durante el último año, y así evaluar el impacto que pudiera tener el comportamiento del tipo de cambio con estos puntos de cambio.

#### **Ponencia P25. Manuel Cortés. ITAM**

Pronóstico con modelos de factores dinámicos.

En el presente trabajo se busca presentar el pronóstico de series de tiempo a través de la descomposición en factores para información almacenada en forma de panel de datos. El principal objetivo es el de estudiar el modelo de factores dinámicos y las transformaciones de desestacionalización de los datos o de estabilización de la varianza para estudiar los distintos estimadores y pronósticos que estos generan para una serie de tiempo con información proveniente de un panel de datos. Se realizaron una serie de simulaciones para el estudio de lo siguiente: primero, comprobar los principales supuestos de convergencia del método de factores dinámicos cuando se usa una estructura de independencia para los errores del modelo y cuando no; segundo, el pronóstico generado por el modelo; y tercero, revisando si al aplicar una transformación se pueden mejorar los pronósticos.

#### **Ponencia P26. José Antonio García Ramírez. Coppel**

Pronósticos vía VAR-PLS

En la charla se propone la metodologíaa VAR-PLS para realizar pronósticos conjuntos hasta h pasos hacia adelante. La metodología parte de un Vector Autorregresivo(VAR) cointegrado y realiza el pronóstico por medio de Mínimos Cuadrados Parciales(PLS) utilizando varios rezagos. Posteriormente se construyen intervalos de confianza de los pronósticos por medio de una adaptación para series de tiempo del método Bootstrap.Se ilustra la metodología con un ejercicio empírico con otras 16 series de tiempo macroeconómicas para pronosticar el INPC mexicano. Los resultados son comparados con los pronósticos obtenidos a través de un modelo VAR con el fin de de comparar laprecisión de la metodología propuesta respecto a un modelo usado comúnmente en la literatura.

#### **Ponencia P27. Juan Pablo Hermosillo Becerra. Facultad de Ciencias, UNAM.**

Modelo de Regresión Aditivo con Valores Extremos Generalizado para una respuesta Binaria (BGEVA)

Estoy trabajando con modelos BGEVA (Binary Generalized Extreme Value Additive Model) para estimar una respuesta binomial de un fenómeno que reporta una proporción pequeña de unos, que son los casos de interés. Se pretende realizar mediante una explicación intuitiva como se puede ir generalizando un Modelo lineal hacia su parte aditiva y luego hacia su versión para Valores Extremos. Mediante la presentación de gráficas de las  $\pi(x)$  de cada modelo se irá mostrando cómo cada modelo tiene características más adecuadas para la estimación de una respuesta binomial con

pocos unos. También se mostrará una breve simulación de datos en R y la estimación de los 3 modelos para ver las diferencias en cuanto a su ajuste.

**Ponencia P28. Arnoldo Daniel Miranda Fournier. Universidad Autónoma Metropolitana**

Modelación de datos composicionales vía mezclas de distribuciones normales multivariadas

Desde siempre el ser humano se ha enfrentado a procesos que involucran Datos, en consecuencia, la modelación de fenómenos reales se ha convertido en una tarea muy importante. Un conjunto de datos muy particular son los que forman las variables composicionales, los datos composicionales son aquellos que describen las partes de un todo y su espacio muestral natural resulta ser el simplex  $p$ -dimensional. El problema del análisis estadístico de datos composicionales ha sido y es una fuente de preocupación para muchos científicos, puesto que es vasta la frecuencia con que aparecen datos de esta índole en las ciencias aplicadas —ciencias de la tierra (geoquímica, petrología, etc), biología, química, ciencias ambientales, economía, medicina, sociología, ingeniería— y, por ende, la importancia de disponer de herramientas adecuadas para su análisis. La metodología que se desarrollo se basa en la propuesta de un modelo no-paramétrico (mezcla infinita de densidades normales multivariadas) para describir datos composicionales.

**Ponencia P29. Lizbeth Naranjo Albarrán. Facultad de Ciencias, UNAM**

Un modelo de Markov oculto para respuesta ordinal con clasificación incorrecta en un proceso no decreciente.

Algunas veces el proceso de generación y recolección de datos es imperfecto, por tanto, se obtienen datos con errores de medición. La medición incorrecta da lugar a una pérdida de información efectiva y a una distorsión de la realidad. Cuando los datos están sujetos a errores de medición, los modelos estadísticos en los que no se tiene en cuenta este hecho producen errores en las estimaciones de los parámetros, tanto en el valor estimado como en su precisión. Se presentará un modelo de Markov oculto para datos con respuesta ordinal con clasificación incorrecta en un proceso no decreciente.

**Ponencia P30. Irving Aarón Díaz Espinoza. Universidad Autónoma de Tlaxcala**

Suavización exponencial doble para predecir los casos confirmados acumulados y defunciones por Covid-19 en México

Este trabajo nos muestra el avance que se ha tenido respecto al número de casos confirmados acumulados y defunciones por Covid-19 en todo el país de México, desde el 27 de febrero hasta la actualidad. Asimismo, se realiza una predicción para los próximos días usando una suavización exponencial doble analizando su efectividad de pronóstico en comparación con la realidad de los casos presentes en México.

**Ponencia P31. Sergio Juárez. Universidad Veracruzana**

Un modelo SEIRD espacial para la epidemia de COVID-19 en el estado de Veracruz

La epidemia de COVID-19 en el estado de Veracruz ha tenido una evolución heterogénea tanto en la mayoría de los municipios con baja incidencia cómo en los focos rojos ubicados en seis zonas

urbanas: Veracruz-Boca del Río, Coatzacoalcos, Orizaba, Córdoba, Xalapa y Poza Rica. En este trabajo presentamos un modelo Susceptibles-Expuestos-Infectados-Recuperados-Defunciones que toma en cuenta la heterogeneidad de estas regiones para estimar las tasas de transmisión del COVID-19. El ajuste del modelo lo hacemos bajo el paradigma de la inferencia Bayesiana y usando los datos de la parte visible de la epidemia: casos oficialmente confirmados y defunciones oficialmente confirmadas.

**Ponencia P32. Arturo Erdely. FES Acatlán, UNAM.**

Midiendo el pulso de una epidemia

Se presentarán características básicas de los datos abiertos de la Secretaría de Salud sobre la epidemia por COVID-19 en México, así como un primer análisis e interpretación de los mismos.

**Ponencia P33. Michelle Anzarut y Teresa Ortiz**

Estimando los casos diarios de COVID-19 y el número de reproducción en México

En este trabajo ajustamos un modelo jerárquico bayesiano semi-mecanicista para describir la epidemia de COVID-19 en México. Obtenemos dos medidas epidemiológicas: el número de infecciones y el número de reproducción. Las estimaciones se basan en datos de muerte, por lo tanto, esperamos que nuestras estimaciones sean más precisas que aquellas que se hacen a partir del número de casos informados.

**Ponencia P34. Jorge Esteban Herrera Plascencia. CIMAT - Aguascalientes**

Evaluación de la Práctica Docente Universitaria Mediante Análisis de Factores

La evaluación docente es la valoración del trabajo del profesor, permitiendo una autoevaluación de las fortalezas y debilidades para la mejora de la práctica docente. No obstante, no todas las instituciones cuentan con instrumentos objetivos para la evaluación, por lo que el objetivo fue valorar la práctica docente a nivel universitario en Los Mochis, Sinaloa. Se diseñó un cuestionario con 28 reactivos para evaluar el grado de importancia de las características y habilidades del docente mediante una escala ordinal de Likert entre 1 y 10, aplicado a 90 estudiantes de licenciatura. Los resultados mostraron que el coeficiente de confiabilidad para la escala total es confiable (alfa de Cronbach = 0.957), y un índice de KMO de 0.93. Después de un análisis factorial con un método de extracción por componentes principales, y una rotación VARIMAX, se eligieron 3 factores que representan el 67.2% de la variabilidad, etiquetados como: 1) competencias instrumentales docentes; 2) competencias interpersonales/interacción; 3) competencias de investigación científica. Posteriormente, 20 docentes evaluados, mediante el instrumento, fueron clasificados con mayores puntajes hacia las relaciones interpersonales; medio para las habilidades instrumentales y de investigación científica.

**Ponencia P35. Leonardo Ailines Genis. CIMAT**

El rezago educativo entre los adolescentes de 12 a 15 años, ¿avance o retroceso en las políticas educativas a nivel mundial?

El presente trabajo de investigación tiene como objetivo estimar los niveles y la tendencia del rezago educativo en los adolescentes de 12 a 15 años a nivel mundial. Para ello se considera el porcentaje de adolescentes que no asisten a la escuela, la cual se define como la proporción de adolescentes en edad de asistir a la secundaria (o equivalente) que nunca han asistido a la escuela o que abandonaron sus estudios. El resultado de este análisis muestra parte el desempeño de las políticas educativas en la reducción del rezago educativo. Asimismo, bajo el modelo logístico, se puede pronosticar si se llegará a cumplir de una de las metas del cuarto objetivo de los ODS: en 2030 asegurar que todas las niñas y niños terminen la enseñanza primaria y secundaria.

### **Ponencia P36. Diana Barraza Barraza. Universidad Juárez del Estado de Durango**

Cuestionario para evaluación de un Programa Institucional de Tutorías: R y Shiny como herramienta para la recolección masiva de datos

En el año 2019, la Red Regional Noreste de Tutorías (RRNET) de la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES) solicitó a la Universidad Juárez del Estado de Durango, a través de la Facultad de Ciencias Exactas, la aplicación y análisis de cuestionarios para evaluar un Programa Institucional de Tutorías (PIT) en 8 Instituciones de Educación Superior ubicadas en siete estados de la República Mexicana. Ante esta necesidad, la Facultad de Ciencias Exactas, decide hacer uso de herramientas estadísticas y computacionales para la aplicación de dichos cuestionarios a gran escala. En esta ponencia se presenta como el uso del lenguaje de programación R y la paquetería Shiny permitió recolectar información de más de 12,000 individuos en un lapso pequeño con mínimos costos operacionales.

### **Ponencia P37. Sergio Rodríguez. Universidad Juárez del Estado de Durango**

Diseño de Experimentos y Regresión Logística como auxiliares para desarrollar metodologías de trabajo colaborativo que desarrollen habilidades algebraicas en bachillerato.

Este trabajo presenta la aplicación de diseño de experimentos y regresión logística multinomial para el desarrollo de una metodología de trabajo colaborativo que permita el desarrollo de habilidades algebraicas en alumnos de bachillerato. Las variables analizadas fueron el estilo de aprendizaje de los alumnos, , acordes al modelo de Honey y Mumford, la especialidad a la que correspondía el grupo, las habilidades algebraicas y el nivel de dificultad de éstas. La propuesta trabajó con 68 alumnos bachilleres distribuidos en tres grupos por carrera técnica, evaluando en tres niveles de complejidad tres habilidades algebraicas - sustitución, jerarquía de las operaciones y despeje de variables. Se usó un diseño factorial completo, para el diseño experimental, y con los resultados, se ajustó un modelo de regresión logística multinomial, con factores anidados, a fin de estimar la probabilidad de que los alumnos mejoraran o no las habilidades estudiadas. En esta ponencia se discutirán las dificultades encontradas para llevar a cabo el diseño de experimentos y el ajuste de modelos de regresión logística, así como resultados preliminares del estudio.