

JEDE_V.irtUAL_2021

**I Encuentro virtual en Diseño Óptimo de
Experimentos**

20, 21 y 22 de enero de 2021

PROGRAMA Y RESÚMENES



**UNIVERSIDAD
DE ALMERÍA**

PROGRAMA

Miércoles, 20 de enero de 2021

Hora	Ponente	Universidad/Institución	Título del trabajo
9:00	Inauguración a cargo del Rector de la Universidad de Almería, Carmelo Rodríguez Torreblanca		
9:30	Isabel María Ortiz	Universidad de Almería	Diseño óptimo de experimentos para modelos de regresión: el diseño del logo de JEDE V
10:00	Edgar Benitez	Universidad de Navarra	Diseño óptimo de modelos de crecimiento tumoral
10:30	Pablo Pérez Sánchez	Universidad Internacional Menéndez Pelayo	Predicción de eventos cardiovasculares y hemorrágicos en pacientes con doble antiagregación con modelos Machine Learning
11:00	Laura del Campo-Albendea	IRYCIS, Hospital Ramón y Cajal	Methodological framework for the development and validation of a prognostic model of mortality in the ICU and in mechanically ventilated patients with ARDS associated with COVID-19
11:30	Helena Fidalgo	Spanish National Cancer Research Center & CIBERONC	Study of the distribution and behavior of the "0" values in large omic datasets

Moderadora: Isabel María Ortiz Rodríguez

Jueves, 21 de enero de 2021

Hora	Ponente	Universidad/Institución	Título del trabajo
9:30	Álvaro Cía Mina	Universidad de Navarra	Active learning: aplicación del diseño óptimo de experimentos en Big Data
10:00	Àngela Sebastià Bargues	Universidad de Castilla-La Mancha	Diseños D-óptimos para modelos de impedancia eléctrica
10:30	Carlos de la Calle Arroyo	Universidad de Castilla La Mancha	Diseño de Experimentos con R
11:00	Pablo Urruchi	Universidad de Navarra	ODE for stock prediction based on Twitter
11:30	Adriana Villa Murillo	Universidad Viña del Mar (Chile)	Los métodos Forest-Genetic para el Diseño Robusto de Parámetros

Moderador: Jesús López Fidalgo

Viernes, 22 de enero de 2021

Hora	Ponente	Universidad/Institución	Título del trabajo
9:30	Juan M. Rodríguez Díaz	Universidad de Salamanca	Optimal Designs for Time-Depending Multiresponse-Multisubject Models
10:00	Sergio Pozuelo Campos	Universidad de Castilla-La Mancha	Effect of misspecification in probability distribution of the response variable in optimal experimental design
10:30	Irene García-Camacha Gutiérrez	Universidad de Castilla-La Mancha	Nuevos avances en la construcción de diseños óptimo-robustos para modelos de tiempo de fallo acelerado en observaciones censuradas por la derecha
11:00	Irene Mariñas-Collado	Universidad de Oviedo	Designs for Accelerated Failure Time models with temperature as acceleration factor
11:30	Clausura del Encuentro		

Moderador: Ignacio Martínez López

I Encuentro virtual en Diseño Óptimo de Experimentos

Miércoles, 20 de enero

Título del trabajo: **Diseño óptimo de experimentos para modelos de regresión: el diseño del logo de JEDE V**

Autores: ***Ignacio Martínez López e Isabel María Ortiz Rodríguez***

Ponente: **Isabel María Ortiz Rodríguez**

RESUMEN

Se hará una introducción al diseño óptimo de experimentos para modelos de regresión, centrándonos en los principales elementos que intervienen en los diseños: modelo, espacio de diseño, matriz de información y criterio de optimización, además de los teoremas de equivalencia para comprobar la optimalidad de un diseño.

Como caso particular se verá el diseño de experimentos que se recoge en el logo del 5º Congreso de Jóvenes Investigadores en Diseño de Experimentos y Bioestadística (JEDE V).

Referencias

1. Atkinson, A.C. y Donev, A.N. *Optimal Experimental Designs*. Oxford Science Publications, 1992.
2. Pukelsheim, F. *Optimal Design of Experiments*. Society for Industrial and Applied Mathematics, 2006.
3. Rodríguez, C. y Ortiz, I. *Diseño óptimo de experimentos para modelos de regresión*. Servicio de Publicaciones de la Universidad de Almería, 1999.

I Encuentro virtual en Diseño Óptimo de Experimentos

Miércoles, 20 de enero

Título del trabajo: **Diseño óptimo de modelos de crecimiento tumoral**

Autores: ***Maite Aznárez- Sanado Maite, Edgar Benítez Edgar y Jesús López-Fidalgo***

Ponente: **Edgar Benítez**

RESUMEN

La evaluación del crecimiento de células tumorales es un aspecto muy importante dentro del estudio clínico de estas patologías. Para tal fin los resultados de investigación han optado por diferentes estrategias en modelización: desde modelos exclusivamente empíricos a modelos funcionales con un alto valor explicativo. En este trabajo se evalúan dos modelos representativos de ambos enfoques y se busca identificar el diseño experimental óptimo bajo criterios de D-optimización, el modelo Gompertz simplificado y un modelo diferencial. Para el primero se muestran los resultados de los puntos óptimos de evaluación y para el modelo diferencial se concluye que la forma explícita no es posible resolverla en el conjunto de los números reales.

Referencias

1. Simeoni, M., Magni, P., Cammia, C., De Nicolao, G., Croci, V., Pesenti, E., & Rocchetti, M. (2004). Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer research*, 64(3), 1094-1101.
2. Vaghi, C., Rodallec, A., Fanciullino, R., Ciccolini, J., Mochel, J. P., Mastri, M., & Benzekry, S. (2020). Population modeling of tumor growth curves and the reduced Gompertz model improve prediction of the age of experimental tumors. *PLoS computational biology*, 16(2), e1007178.

I Encuentro virtual en Diseño Óptimo de Experimentos

Miércoles, 20 de enero

Título del trabajo: **Predicción de eventos cardiovasculares y hemorrágicos en pacientes con doble antiagregación con modelos Machine Learning**

Autores: ***Pablo Pérez Sánchez, Antonio Sánchez Puente, Pedro Luis Sánchez Fernández y José Manuel Sánchez Santos***

Ponente: **Pablo Pérez Sánchez**

RESUMEN

El infarto agudo de miocardio es una de las enfermedades con mayor impacto en España. Aunque el intervencionismo percutáneo temprano tiene altas tasas de éxito para tratar la enfermedad, los pacientes se enfrentan a continuación al riesgo de reinfarto y sangrado. Una de las terapias utilizadas para disminuir esta incidencia es la Doble Antiagregación Plaquetaria. Durante esta charla se expondrán las técnicas de estadística e Inteligencia Artificial utilizadas para la predicción del riesgo de reinfarto y sangrado de pacientes que han seguido la terapia doble antiagregación.

Referencias

1. Annual Incidence of Confirmed Stent Thrombosis and Clinical Predictors in Patients With ACS Treated With Ticagrelor or Prasugrel. Rev Esp Cardiol (Engl Ed). 2019;72(4):298-304. (Abril 2019).
2. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer. (2009).
3. Artificial Intelligence: A Modern Approach. Prentice Hall. (2010)

I Encuentro virtual en Diseño Óptimo de Experimentos

Miércoles, 20 de enero

Título del trabajo: **Methodological framework for the development and validation of a prognostic model of mortality in the ICU and in mechanically ventilated patients with ARDS associated with COVID-19**

Autores: ***L. del Campo-Albendea, A. Muriel-García, J. Zamora, F. Frutos-Vivar y O. Peñuelas***

Ponente: **Laura del Campo Albendea**

RESUMEN

The objective was developed a prognostic model of mortality in the ICU for patients with COVID-19 requiring invasive ventilation. The ICU mortality variable was analyzed using a mixed-effects logistic model, whit patients nested in the different ICUs. We applied stepwise selection with backward elimination and missing data were imputed using the MICE methodology. The final model was evaluated by bootstrapping.

A total of 868 patients were included. The predictive factors associated with ICU mortality were: age (OR 1.049 [95%CI 1.032-1.066]), SAPS3 score (OR 1.025 [95%CI 1.008-1.041]), neutrophils/lymphocytes ratio (OR 1.009 [95%CI 1.002-1.016]), use of non-invasive ventilation prior to ICU admission (OR 2.131 [95%CI 1.279-3.550]), use of selective digestive decontamination (OR 0.587 [95%CI 0.358-0.963]) and need for low-dose vasopressor support (OR 1.878 [95%CI 1.119-3.152]). These coefficients were used to generate a nomogram.

Reporting validation results of the predictive models of ICU mortality developed based on advanced statistical techniques, allows the ability to generalize the results obtained and providing a predictive tool valuable for clinical utility.

Referencias

1. Gupta S, Hayek SS, Wang W, et al. Factors Associated With Death in Critically Ill Patients With Coronavirus Disease 2019 in the US. JAMA Intern Med. Published online July 15, 2020, e203596.
2. Miller ME, Langefeld CD, Tierney WM, Hui SL, McDONALD CJ. Validation of probabilistic predictions. Med Decis Making 1993; 13: 49-58.

I Encuentro virtual en Diseño Óptimo de Experimentos

Miércoles, 20 de enero

Título del trabajo: **Study of the distribution and behavior of the “0” values in large omic datasets**

Autores: **Helena Fidalgo, Teresa Pérez, Lola Alonso y Núria Malats**

Ponente: **Helena Fidalgo Gómez**

RESUMEN

Microbiome datasets are characterized by a low number of positive counts and an excess of zero (EoZ), resulting in high overdispersion. We propose a methodology that allows us to model the taxon-count data, and treat the EoZ as an advantage: the Zero-Inflated Models and Zero-Hurdle Models [1]. The distributions assumed in both models can differentiate between 3 types of zeros that are in the microbiome datasets: outliers, structural, and sampling [2]. Both models and their distributions, are evaluated in several scenarios with different percentages of zeros and using various model selection criteria [3]. In addition, a graphical diagnostic has been performed, including the Q-Q Plots and the Rootogram. We found that the Zero-Hurdle Model is the model that best adapts to the microbiome-count data. It, has made possible to automate build a workflow for large taxon-counting databases with EoZ.

All these methods have been tested using a real dataset included in The Cancer Genome Atlas, in particular those related with the Urothelial Bladder Carcinoma. We analyze the relationships between certain bacteria and several covariates.

Referencias

1. Kaul, A., Mandal, S., Davidov, O., & Peddada, S. D. (2017). Analysis of Microbiome Data in the Presence of Excess Zeros. *Frontiers in microbiology*, 8, 2114. <https://doi.org/10.3389/fmicb.2017.02114>
2. Xia, Y., Sun, J., & Chen, D.-G. (2018). Introductory Overview of Statistical Analysis of Microbiome Data. In Y. Xia, J. Sun, & D.-G. Chen (Eds.), *Statistical Analysis of Microbiome Data* (pp. 43–75). Springer. https://doi.org/10.1007/978-981-13-1534-3_3
3. Xu, L., Paterson, A. D., Turpin, W., & Xu, W. (2015). Assessment and Selection of Competing Models for Zero-Inflated Microbiome Data. *PLOS ONE*, 10(7), e0129606. <https://doi.org/10.1371/journal.pone.0129606>

I Encuentro virtual en Diseño Óptimo de Experimentos

Jueves, 21 de enero

Título del trabajo: **Active learning: aplicación del diseño óptimo de experimentos en Big Data**

Autores: *Álvaro Cía Mina y Jesús F. López Fidalgo*

Ponente: **Álvaro Cía Mina**

RESUMEN

Una de las aplicaciones del diseño óptimo de experimentos en Big Data son los métodos de active learning para submuestreo óptimo de grandes muestras, los cuales nos proporcionan una mejor estimación de los parámetros de la población. Nuestro trabajo se centra en el estudio de la influencia de la distribución de las variables explicativas en estos métodos.

I Encuentro virtual en Diseño Óptimo de Experimentos

Jueves, 21 de enero

Título del trabajo: **Diseños D-óptimos para modelos de impedancia eléctrica**

Autores: ***Raúl Martín Martín, José Luis Polo Sanz y Àngela Sebastià BARGUES***

Ponente: **Àngela Sebastià BARGUES**

RESUMEN

El comportamiento eléctrico de una gran variedad de sistemas se puede caracterizar por medidas de impedancia eléctrica [1,2], es decir, la oposición que presenta un sistema específico en estudio al flujo de corriente eléctrica en función de la frecuencia. El modelo matemático utilizado es un circuito de Randles simplificado, en el que la combinación en paralelo de un condensador y una resistencia describe la impedancia de la forma más simple de una interfaz de electrodo corrosivo o el comportamiento pasivo eléctrico de una membrana celular. Formulamos el problema de estimar parámetros en el sistema lineal de entrada única y salida única (SISO) en el dominio de la frecuencia utilizando la teoría del análisis de diseño óptimo de experimentos [3]. Para obtener las señales de entrada óptimas para la estimación de parámetros, calculamos diseños óptimos aproximados y exactos optimizando el determinante de la matriz de información adaptando dos de los algoritmos que más se utilizan habitualmente en la actualidad, el algoritmo de intercambio aleatorio REX y el algoritmo de intercambio KL.

Referencias

1. E. Barsoukov and J. R. Macdonald. Impedance Spectroscopy: Theory, Experiment and Applications. 3rd ed. New Jersey: Wiley. (2018)
2. Graham C. Goodwin, R. L. Payne. Dynamic System Identifications: Experiment design and data analysis. Mathematics in science and engineering: Volume 136. Academic Press. (1977)
3. A. C. Atkinson, A. N. Donev and R. D. Tobias. Optimum Experimental Designs with SAS. Oxford: University Press. (2007)

I Encuentro virtual en Diseño Óptimo de Experimentos

Jueves, 21 de enero

Título del trabajo: **Diseño de Experimentos con R**

Autores: ***Carlos de la Calle-Arroyo, Jesús López-Fidalgo y Licesio J. Rodríguez-Aragón***

Ponente: **Carlos de la Calle Arroyo**

RESUMEN

El cálculo de diseños óptimos es un problema complejo para el que a menudo no se puede encontrar una solución analítica. Por lo tanto, se suele recurrir a la utilización de algoritmos, implementados en distintos tipos de software.

En esta presentación vamos a comentar algunas de las opciones disponibles en R para el cálculo de diseños óptimos, así como introducir el trabajo que estamos desarrollando en este sentido, utilizando las ventajas que ofrece R.

Referencias

1. Filová L, Harman R (2020): Ascent with quadratic assistance for the construction of exact experimental designs, Computational Statistics, Volume 35, pp. 775–801
2. Masoudia E., Hollinga H., Wong W.K. (2017): Application of imperialist competitive algorithm to find minimax and standardized maximin optimal designs, Computational Statistics & Data Analysis, Volume 113, pp. 330-345

I Encuentro virtual en Diseño Óptimo de Experimentos

Jueves, 21 de enero

Título del trabajo: **ODE for stock prediction based on Twitter**

Autores: ***Pablo Urruchi Mohino y Jesús López Fidalgo***

Ponente: **Pablo Urruchi Mohino**

RESUMEN

We have designed a posterior distribution to be used as a penalization in maximum a posteriori optimization problem following a Bayesian approach. Our goal in optimal design of experiments is to use this framework to be included in an existing model which tries to predict a stock movement based on Twitter comments for that particular stock, selecting the most relevant features within tweets text.

I Encuentro virtual en Diseño Óptimo de Experimentos

Jueves, 21 de enero

Título del trabajo: **Los métodos Forest-Genetic para el Diseño Robusto de Parámetros**

Autora: **Adriana Villa Murillo**

Ponente: **Adriana Villa Murillo**

RESUMEN

La ingeniería de calidad surge como una disciplina para detectar y prevenir problemas de calidad desde las etapas tempranas del desarrollo y diseño del producto, hasta los problemas asociados con sus funciones y costes derivados de la fabricación y puesta en el mercado. En los años 80, Taguchi propuso enfatizar el esfuerzo en la calidad del producto desde el diseño del mismo, desarrollando una aproximación al diseño de experimentos. Esto se conoce como Diseño Robusto de Parámetros.

Tales ideas causaron controversias desde sus inicios, dando origen a diversas alternativas donde se incorporan herramientas de inteligencia artificial. Los métodos Forest-Genetic es uno de los más recientes, donde se combina a Random Forest y los Algoritmos Genéticos en 3 fases: normalización, modelización y optimización; para problemas de tipo univariante y multivariante.

Dados los buenos resultados numéricos obtenidos hasta ahora, en comparación con otras propuestas, la presente comunicación tiene como objetivo compartir las bases de tales metodologías y se muestra brevemente las aplicaciones que actualmente se están llevando a cabo en controles fitosanitarios.

Referencias

1. Adriana Villa-Murillo, Andrés Carrión García, Antonio Sozzi. Forest-Genetic method to optimize parameter design of multiresponse experiment. *Inteligencia Artificial*. 23(66). 2000. 9-25. doi: 10.4114/intartif.vol23iss66pp9-2

I Encuentro virtual en Diseño Óptimo de Experimentos

Viernes, 22 de enero

Título del trabajo: **Optimal Designs for Time-Depending Multiresponse-Multisubject Models**

Autores: **Juan M. Rodríguez Díaz, Rosa Eva Pruneda y M. Mercedes Rodríguez-Hernández**

Ponente: **Juan M. Rodríguez Díaz**

RESUMEN

The time-evolution of characteristics from a population that may be correlated will be studied from the point of view of design. Depending on the type of the observations, the corresponding tests needed for the follow-up of the variables could be expensive or time-demanding, thus usually the aim is to decide the most convenient moments where a limited number of tests should be done in order to obtain the maximum information. Using an optimal design of experiments approach, a general method for selection of the best temporal points will be developed.

An application to the study of two variables related with the capacity of resolution of mathematical problems for primary-school students will be shown, employing different evolution models.

Referencias

1. Rodríguez Díaz J.M. and Sánchez-León G. (2019a). Efficient parameter estimation in multiresponse models measuring radioactivity retention. *Radiation and Environmental Biophysics* 58 (2): 167--182
2. Rodríguez Díaz J.M. and Sánchez-León G. (2019b). Optimal designs for multiresponse models with double covariance structure. *Chemometrics and Intelligent Laboratory Systems*, 189 1-7
3. Rodríguez-Hernández M., Pruneda R.E. and Rodríguez Díaz J.M. (2020). Statistical Analysis of the Evolutionary Effects of Language Development in the Resolution of Mathematical Problems in Primary School Education. Technical Report 23, Statistics Department, University of Salamanca

I Encuentro virtual en Diseño Óptimo de Experimentos

Viernes, 22 de enero

Título del trabajo: **Effect of misspecification in probability distribution of the response variable in optimal experimental design**

Autores: ***Sergio Pozuelo Campos, Mariano Amo Salas y Víctor M. Casero Alonso***

Ponente: **Sergio Pozuelo Campos**

RESUMEN

It is a common assumption in the context of optimal experimental design that the response variable follows a homoscedastic normal distribution. There are, however, other studies that assume different probability distributions based on prior experience or additional information. The main goal of this study is to study the effect, in terms of efficiency, of misspecification in the probability distribution on optimal design. From the elemental information matrix, which includes information on the probability distribution of the response variable, a generalized Fisher information matrix is derived. Relevant theoretical results were obtained, for different regression models, comparing heteroscedastic normal with gamma or Poisson distributions. Finally, the analysis was broadened to include a practical case which considers a 4-parameter Hill model, to explain the effect of several drugs on cell development.

Referencias

1. Atkinson, A. C., Fedorov, V. V., Herzberg, A. M., and Zang, R. (2014). Elemental information matrices and optimal experimental design for generalized regression models. *Journal of Statistical Planning and Inference*, 144:81–91.
2. Khinkis, L. A., Levasseur, L., Faessel, H., and Greco, W. R. (2003). Optimal design for estimating parameters of the 4-parameter Hill model. *Nonlinearity in Biology, Toxicology and Medicine*, 1:363–377.
3. McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall/CRC.

I Encuentro virtual en Diseño Óptimo de Experimentos

Viernes, 22 de enero

Título del trabajo: **Nuevos avances en la construcción de diseños óptimo-robustos para modelos de tiempo de fallo acelerado en observaciones censuradas por la derecha**

Autores: ***María Jesús Rivas López, Raúl Martín Martín e Irene García-Camacha Gutiérrez***

Ponente: **Irene García-Camacha Gutiérrez**

RESUMEN

Los modelos de tiempo de fallo acelerado (AFT) se definen a través de la función de supervivencia del tiempo hasta un evento. Este trabajo persigue construir diseños óptimo-robustos para modelos AFT con la posibilidad de que el factor de aceleración esté especificado parcialmente a través de una función contaminante. En este caso, la estimación de los parámetros del modelo y de la respuesta predicha está sujeta no solo al error en la varianza debido al muestreo, sino también al error de sesgo producido por la falta de adecuación del modelo. Se ha hallado la matriz del error cuadrático medio asintótica de los estimadores para observaciones censuradas por la derecha. Finalmente, se presenta una aplicación de los resultados para el caso del modelo AFT log-logístico en el que se calculan los diseños D- e I-óptimo robustos.

I Encuentro virtual en Diseño Óptimo de Experimentos

Viernes, 22 de enero

Título del trabajo: **Designs for Accelerated Failure Time models with temperature as acceleration factor**

Autores: ***Irene Mariñas-Collado, M. Jesús Rivas-López, Juan M. Rodríguez-Díaz y M. Teresa Santos-Martín***

Ponente: **Irene Mariñas del Collado**

RESUMEN

An accelerated life test of a product or material consists on the observation of its failure time when it is subjected to conditions that stress the usual ones. The purpose is to obtain the parameters of the distribution of the time-to-failure for usual conditions through the observed failure-times. Increasing temperature is one of the most commonly used methods to accelerate a failure mechanism. Different optimal plans for AFT models, when the accelerated failure process is described making use of Arrhenius or Eyring equations, are studied.