

# III JORNADAS DE ESTADÍSTICA COMO HERRAMIENTA CIENTÍFICA

DEL 5 AL 7  
FEBRERO 2020



<https://estadisticacongres.wixsite.com/inicio>



Universidad de Jaén

Departamento de Estadística e  
Investigación Operativa y Escuela de Doctorado.



## ÍNDICE

<i>Determinants of Overfunding in Equity Crowdfunding: An Empirical Study</i> .....	2
<i>Análisis estadístico multivariante para el estudio de la endometriosis y sus enfermedades comórbidas</i> .....	5
<i>Diferencias en la gestión financiera en el fútbol español. Ligas Europeas vs Liga Nacional</i> .....	7
<i>Detección de denso-dependencia en ungulados cinegéticos mediante un modelo estadístico basado en el filtro de Kalman</i> .....	10
<i>Aplicación de la Estadística en la Evaluación de Competencias Escritoras</i> .....	11
<i>Calidad de vida en personas mayores con síndrome de apnea del sueño. Relación con el deterioro cognitivo, deterioro funcional y enfermedades crónicas</i> .....	13
<i>Encefalopatía Epiléptica STXBP1: predictores computacionales y estadísticos de asociación genotipo/fenotipo</i> .....	15
<i>Estudio de la validez de la Autoevaluación como método de evaluación educativa: uso del bootstrapping</i> .....	18
<i>Regresión Logística Multinivel Multinomial aplicada a estudios en turismo</i> .....	20
<i>Alfabetización estadística y fake news, una herramienta contra la desinformación</i> .	22
<i>Análisis de alteraciones neurológicas complejas mediante teoría de juegos cooperativos aplicada a datos de microarrays</i> .....	24
<i>La microestructura en los libros de la enseñanza secundaria obligatoria: las medidas de dispersión</i> .....	27
<i>Indicadores de transparencia para organizaciones sin ánimo de lucro: Estudio de un caso español</i> .....	28
<i>La estadística como herramienta de búsqueda de genes candidatos en los trastornos del espectro autista</i> .....	31
<i>Asociación entre genotipo y respuesta farmacológica en la encefalopatía epiléptica STXBP1</i> .....	36
<i>Tratamiento de la multicolinealidad en el modelo CAMELS a través de la regresión Cresta</i> .....	39
<i>Valor de la transparencia en entidades sin ánimo de lucro: propuesta de mejora de una batería de indicadores</i> .....	40
<i>Simulación de un modelo de regresión de Poisson para datos de panel</i> .....	42
<i>Lectura de tablas estadísticas por futuras maestras de Educación Infantil en Chile</i> ..	43
<i>Estudio de ciertas falacias cometidas en los medios de comunicación</i> .....	44
<i>Comprensión de la selección de muestras</i> .....	46

<i>Algunas implicaciones cívicas de la estadística: El caso del proceso electoral en España .....</i>	<i>48</i>
<i>Propuesta de un programa de intervención sobre problemas adaptativos asociados a la dependencia emocional .....</i>	<i>51</i>
<i>La Probabilidad en la enseñanza.....</i>	<i>52</i>
<i>Una modelización para los centros de educación .....</i>	<i>53</i>



# Determinants of Overfunding in Equity Crowdfunding: An Empirical Study

Cristina Martínez Gómez<sup>(1)</sup>, Francisca Jiménez Jiménez<sup>(2)</sup>, M<sup>a</sup> Virtudes Alba Fernández<sup>(3)</sup>

(1) *Departamento de Economía, Universidad de Jaén, cmgomez@ujaen.es*

(2) *Departamento de Economía, Universidad de Jaén, jimenezf@ujaen.es*

(3) *Departamento de Estadística e Investigación Operativa, Universidad de Jaén, mvalba@ujaen.es*

## RESUMEN

Esta investigación pretende identificar el impacto de factores influyentes en el nivel de éxito que consiguen los proyectos de crowdfunding de inversión. El crowdfunding ofrece a las empresas una solución a los problemas de financiación que se enfrentan, en especial las pequeñas empresas que necesitan un empuje financiero en sus etapas iniciales. La expansión del uso de esta herramienta da lugar a su auge como alternativa de financiación. Este estudio muestra una mayor comprensión de los factores que conducen a un mayor éxito en los proyectos de crowdfunding, pudiendo alcanzar diferentes niveles de éxito, motivado por las características del mismo, algunos de ellos situándose muy por encima de la media de la financiación total alcanzada.

Para medir el nivel de éxito consideramos la sobrefinanciación, cantidad de financiación alcanzada y número de inversores como variables clave. Los factores que influyen en el nivel de éxito se han agrupado en cuatro categorías: el nivel de riesgo, las características de calidad del proyecto, la educación de los fundadores y los incentivos fiscales de los inversores. La información procede de una muestra de 344 proyectos anunciados en la plataforma Crowdcube en España y UK, entre abril de 2015 y julio de 2018.

Un primer estudio exploratorio da lugar a diferencias significativas en la distribución de las variables que explican el nivel de éxito según el país (proyectos desarrollados en España o en UK). Posteriormente, se analiza el impacto de los factores determinantes por país. Los resultados econométricos basados en regresiones no paramétricas (regresiones cuantiles) muestran que los efectos significativos de ciertos factores dependen del valor de los distintos cuantiles. Factores como la presencia de las empresas en las redes sociales pierden importancia para proyectos con una sobrefinanciación y una cantidad de financiación alcanzada ubicados ambos en el cuartil 75. Asimismo, el factor equity tiene un impacto significativo para proyectos con una sobrefinanciación ubicada entre los cuantiles 25 y 50 y una cantidad de financiación en el cuartil 25. Este factor ya no es relevante para los proyectos del 75 cuartil ni para la variable número de inversores. Por otro lado, hay factores que se vuelven determinantes solo para el cuartil 50 y 75 para explicar el número de inversores, como es el caso del derecho al voto. Otro factor determinante del éxito en la campaña de financiación es el tiempo que la empresa lleva operando en su sector. El análisis muestra que el tiempo de experiencia es relevante para el número de inversores en el cuartil 75. Sin embargo, para la sobrefinanciación solo resulta significativo en el cuartil 25.

Por lo tanto, los resultados de la presente investigación demuestran un comportamiento claramente diferenciador entre los proyectos que logran un alto nivel de éxito y aquellos otros que están cerca del objetivo establecido por los fundadores. Además, la relevancia de ciertos factores difiere en algunos casos para el análisis de los proyectos que explican el fracaso o éxito de las campañas de crowdfunding de inversión.

**Palabras clave:** equity, overfunding, crowdfunding, success.

## ABSTRACT

This research will allow us to identify the impact of influential factors on the level of success of a project funded through an equity crowdfunding campaign. Crowdfunding offers companies a solution to the financing problems faced, especially small businesses. They need a financial push in their initial stages. This study offers a better understanding of the factors that lead to greater success in crowdfunding projects, being able to achieve different levels of success, motivated by the characteristics of the them, some of them being well above the average of the total financing achieved.

To measure the level of success we consider the overfunding, amount of financing achieved and number of investors as variables. Those factors that influence the level of success are grouped into four categories: the risk level, the quality characteristics of the project, the education of the founders and the tax incentives of investors. It was used the information obtained from 345 projects collected from Spain and UK Crowdcube platform, between April 2015 to July 2018.

One of the main differences observed through the descriptive analysis is the difference in the distribution of the variables that explain the level of success between the country (projects developed in Spain or in UK). A non-parametric model is estimated to explain the effect of the variables on the level of success, namely, a quantile regression. The econometric results obtained show significant effects of certain factors, depending on the different quartiles. Certain variables such as the presence of companies in social networks lose importance for projects with an overfunding and an amount of financing achieved located in quartile 75th. Likewise, equity has a significant affect for projects with an overfunding located between quartiles 25th and 50th and a funding amount in quartile 25th. This factor is no longer relevant for 75th quartile projects or for the variable number of investors. On the other hand, there are factors that become decisive only from quartile 50th and 75th for the number of investors, such as the right to vote. Some of the companies that request financing have been in their sector for some time, the variable they collect is characteristic in days of operation, it is relevant for the number of investors in quartile 75th. However, days will only be significant in the quartile 25th for overfunding.

Therefore, the results of the present investigation demonstrate different behaviour between those projects that achieve a high level of success and those that are close to the goal set by

the founders. In addition, a relevance of factors that are in some cases different from factors that explain the failure or success of equity crowdfunding campaigns.

**Keywords:** equity, overfunding, crowdfunding, success.

# *Análisis estadístico multivariante para el estudio de la endometriosis y sus enfermedades comórbidas*

Eva Vargas<sup>(1)</sup>, Signe Altmäe<sup>(2,3,4)</sup>, Francisco J. Esteban<sup>(5)</sup>

(1) Departamento de Biología Experimental, Universidad de Jaén, e-mail: [evargas@ujaen.es](mailto:evargas@ujaen.es)

(2) Departamento de Bioquímica y Biología Molecular I, Universidad de Granada, e-mail: [signealtmae@ugr.es](mailto:signealtmae@ugr.es)

(3) Competence Centre on Health Technologies, Universidad de Tartu, Estonia

(4) Instituto de Investigación Sanitaria ibs, Granada

(5) Departamento de Biología Experimental, Universidad de Jaén, e-mail: [festeban@ujaen.es](mailto:festeban@ujaen.es)

## RESUMEN

La endometriosis es una condición crónica dolorosa que afecta a alrededor de 176 millones de mujeres en todo el mundo y que se caracteriza por la presencia de tejido endometrial fuera de la cavidad uterina. Si bien se desconoce la causa por la que se produce, se ha descrito su comorbilidad con otras patologías como las enfermedades autoinmunes, distintos tipos de cáncer o alergias. El objetivo de este trabajo fue detectar la existencia de un patrón de expresión génica común entre la endometriosis y sus enfermedades comórbidas que pudiera explicar su co-ocurrencia. Para ello, se realizó una revisión sistemática con el objetivo de identificar enfermedades comórbidas a la endometriosis a partir de la literatura disponible en la base de datos Pubmed, y posteriormente se analizó el perfil de expresión de las mismas utilizando una estrategia propia de la biología de sistemas, a través de la construcción y análisis de dendrogramas, redes de interacción génica y análisis de enriquecimiento funcional. El análisis estadístico de los datos incluyó técnicas de remuestreo y estimaciones de error mediante *bootstrap*. Los resultados obtenidos apuntan a un grupo de 127 genes implicados en rutas biológicas cuya desregulación podría subyacer al desarrollo de la endometriosis y algunas de sus enfermedades comórbidas. Además, mediante la utilización de datasets experimentales independientes, fuimos capaces de validar la expresión diferencial de 16 de estos genes (*AGTR1*, *BDNF*, *C3*, *CCL2*, *CD40*, *CYP17A1*, *ESR1*, *IGF1*, *IGF2*, *IL10*, *MMP1*, *MMP7*, *MMP9*, *PGR*, *SERPINE1* y *TIMP2*) en muestras de tejido endometrial de pacientes con endometriosis. Para algunos de estos genes se ha descrito su participación en la respuesta a fármacos y respuesta inmune, procesos clave para la instauración y desarrollo de la enfermedad. Además, algunos de ellos albergan polimorfismos que predisponen tanto a la endometriosis como a sus comorbilidades. En conclusión, los resultados de nuestro trabajo ponen de manifiesto los procesos moleculares subyacentes a la etiopatogénesis de la endometriosis y sus enfermedades comórbidas, e identifican biomarcadores putativos de endometriosis.

**Palabras clave:** biología de sistemas, comorbilidad, endometriosis, enriquecimiento funcional

## ABSTRACT

Endometriosis is a chronic pain condition that affects around 176 million women worldwide. It is characterised by the presence and growing of endometrial tissue outside the uterine cavity. Although its cause remains still unknown, its comorbidity with other conditions such as autoimmune diseases, cancer or allergies has been widely described. The aim of this work was to identify a common gene expression pattern between endometriosis and its comorbid diseases that could explain their co-occurrence. A systematic review was carried out in order to identify a set of endometriosis comorbid conditions using Pubmed database. Next, the gene expression pattern associated to each disease was analysed using a systems biology approach, namely by using clustering analysis, gene interaction networks and gene set enrichment analysis. Statistical analysis of the data included resampling and bootstrap techniques for error estimation. The current study allowed us to point out that there was a group of 127 genes and their involved biological pathways whose deregulation may underlie the development of endometriosis and some of its comorbidities. In addition, we were able to validate the differential expression of 16 out of these genes (*AGTR1*, *BDNF*, *C3*, *CCL2*, *CD40*, *CYP17A1*, *ESR1*, *IGF1*, *IGF2*, *IL10*, *MMP1*, *MMP7*, *MMP9*, *PGR*, *SERPINE1* and *TIMP2*) in independent experimental datasets of endometrial tissue samples from women with endometriosis. Some of these genes have been described as involved in response to drugs and immune response, which are well-known hallmarks of endometriosis. Moreover, some of these genes harbor polymorphisms that predispose either to endometriosis and its comorbidities. In brief, the results of our study highlight the molecular processes underlying the ethiopathogenesis of endometriosis and its comorbid conditions, and identify putative biomarkers of endometriosis.

**Keywords:** comorbidity, endometriosis, gene set enrichment, systems biology

# *Diferencias en la gestión financiera en el fútbol español. Ligas Europeas vs Liga Nacional*

Alberto Calahorra López <sup>(1)</sup>, Julio Vena Oya <sup>(2)</sup>, Ana Belén Mudarra Fernández <sup>(3)</sup>,  
Melinda Ratkai, PhD <sup>(4)</sup>

(1) Departamento de Organización de Empresas, Marketing y Sociología, Universidad de Jaén, e-mail: [acalahor@ujaen.es](mailto:acalahor@ujaen.es)

(2) Departamento de Organización de Empresas, Marketing y Sociología, Universidad de Jaén, e-mail: [jvena@ujaen.es](mailto:jvena@ujaen.es)

(3) Departamento de Organización de Empresas, Marketing y Sociología, Universidad de Jaén, e-mail: [amudarra@ujaen.es](mailto:amudarra@ujaen.es)

(4) Les Roches Education, Marbella, e-mail: [Melinda.ratkai@lesroches.es](mailto:Melinda.ratkai@lesroches.es)

## RESUMEN

La desigualdad entre los clubes que juegan competiciones europeas y aquéllos que solo participan en ligas nacionales han ido aumentando desde la implantación del Fair Play Financiero (FFP), a pesar de que paralelamente ha ayudado a sanear los estados financieros de los clubes europeos. Usando un método de test paramétrico analizamos si existen diferencias entre las medias de ciertos ratios financieros entre los equipos que obtienen plaza en competiciones europeas en la temporada 2017-2018 y aquéllos que no la obtienen.

## INTRODUCCIÓN

Desde Rottember (1957), El Hodiri & Quirk (1971), Barajas & Rodríguez (2014) y Barget & Chavinier-Rela (2017), se habla de la importancia de una buena gestión en los clubes deportivos.

Tras varios intentos fallidos de reducir los problemas financieros por parte de las administraciones públicas que atravesaban los equipos de las grandes ligas, la entrada del FFP ha conseguido sanear la salud financiera de los clubes europeos (Franck, 2014). Por ello nos hemos planteado comprobar si existen diferencias entre la situación financiera medida con los ratios de endeudamiento, liquidez, ROA y ROE entre los clubes que participarán en competiciones europeas y aquellos que no han conseguido clasificarse durante la temporada 2017-2018 de la liga española.

## RESULTADOS

A través de rSabi se han obtenido los últimos datos financieros pertenecientes a: ratios de liquidez, endeudamiento, ROA y ROE como indicadores de salud financiera en base de las cuentas anuales 30 junio 2019. Hemos utilizado el método de test paramétrico para el estudio a través de SPSS con el fin de determinar la relación existente entre dichos ratios y el objetivo deportivo de participar en competiciones europeas. En la tabla 1 se puede observar los diferentes contrastes sobre la varianza y la media de cada uno de las ratios,

tomando como variable de agrupación el haberse clasificado (1) o no (0) para competición europea. Los clubes clasificados para ligas europeas eran los siguientes: F.C.Barcelona, At. Madrid, Real Madrid, Valencia, Villarreal, Betis, Sevilla. Los clubes no clasificados para competiciones europeas eran: Getafe, Eibar, Girona, Espanyol, Real Sociedad, Celta de Vigo, Alavés, Levante, Athletic Club, Leganés, Deportivo Coruña, Las Palmas, Málaga.

Para ello primero se comprobará la hipótesis de igualdad de varianzas mediante el test de Levene:

$$H_0: \sigma^2_{x_1} = \sigma^2_{x_2}$$

$$H_1: \sigma^2_{x_1} \neq \sigma^2_{x_2}$$

En concreto vemos como solo para el ratio de liquidez podemos asumir que dichas varianzas difieren (a un nivel de significación del 5%), por ello, excepto para dicho ratio, se realiza la prueba t de igualdad de medias asumiendo varianzas iguales:

$$H_0: \mu_{x_1} = \mu_{x_2}$$

$$H_1: \mu_{x_1} \neq \mu_{x_2}$$

Se encuentran diferencias en los ratios de liquidez, el ROA y el ROE, pudiendo afirmarse pues, que en efecto, las medias de los indicadores difieren entre los clubs que alcanzan el objetivo europeo y los que no lo logran.

Tabla 1: diferencias de medias

Ratio	Prueba F de Levene de igualdad de varianzas (p-valor)	Prueba t para igualdad de medias (p-valor)
Liquidez	9.11 (0.01)	17.47 (0.01)
Endeudamiento	0.49 (0.49)	-0.10 (0.92)
ROA	3.47 (0.08)	-2.45 (0.02)
ROE	1.26 (0.27)	-2.11 (0.05)

## CONCLUSIONES

El estudio demuestra que existen diferencias entre los ratios de liquidez, ROA y ROE entre los clubes que obtienen plaza para participar en competiciones europeas y entre los clubes que no lo logran durante la temporada 2017-2018 de la primera división de la liga española. Sin embargo el ratio de endeudamiento no hemos observado diferencias, lo que podría estar en línea con los últimos estudios publicados que indican que desde la implantación del FFP se han reducido las deudas en general de los clubes de las cinco grandes ligas. Sin embargo, como el actual estudio también lo confirma, la diferencia entre los que participan en competiciones europeas y aquéllos que no, se está incrementando.

**Palabras clave:** ratios financieros, competición europea, gestión financiera, clasificación deportiva, soccer.

## BIBLIOGRAFIA

Barajas, A., & Rodríguez, P. (2014). Spanish football in need of financial therapy: Cut expenses and inject capital. *International Journal of Sport Finance*, 9(1), 73-90.

Barget, É., & Chavinier-Rela, S. (2017). An analysis of the diversity of financial resources in grassroots sports clubs: A European perspective. [Analyse de la diversité des recettes des clubs sportifs amateurs: Une perspective Européenne] *Staps*, 116(2), 7-25. doi:10.3917/sta.116.0007.

Franck, E. (2014). Financial fair play in European club football: What is it all about? *International Journal of Sport Finance*, 9(3), 193-217.

Quirk, J. y El Hodiri, M. (1974). *The Economic Theory of a Professional Sports League*. Washington: Brookings Institution.

Rottenberg, S. (1956). The baseball Players' Labor Market. *Journal of Political Economy*, (64), 242-258.

# *Detección de denso-dependencia en ungulados cinegéticos mediante un modelo estadístico basado en el filtro de Kalman*

Antonio Jesús López-Montoya<sup>(1)</sup>, Concepción Azorit<sup>(2)</sup>, Irene García-Garrido<sup>(3)</sup>

(1) Dpto. de Estadística e Investigación Operativa. Universidad de Jaén, España, [amontoya@ujaen.es](mailto:amontoya@ujaen.es)

(2) Dpto. Biología Animal, Vegetal y Ecología, Universidad de Jaén, España, [cazorit@ujaen.es](mailto:cazorit@ujaen.es)

(3) Dpto. de Métodos Cuantitativos para la Economía y la Empresa. Universidad de Granada, España, [irenegarciag@ugr.es](mailto:irenegarciag@ugr.es)

## RESUMEN

La detección de factores influyentes en la dinámica de las poblaciones de ungulados cinegéticos, es de vital importancia para una gestión cinegética apropiada de estas poblaciones. Uno de los factores más influyentes en la dinámica poblacional de ungulados es la denso-dependencia (DD), que actúa como un mecanismo de autorregulación que afecta a las tasas de reclutamiento y de crecimiento poblacional en condiciones de sobreabundancia. Por tanto, detectar su presencia es esencial para considerarla en una gestión cinegética adecuada. En este trabajo se ha utilizado un modelo estadístico basado en el filtro de Kalman para detectar la presencia de DD, y ver cómo actúan las variables climáticas y la caza sobre la dinámica poblacional en dos poblaciones de ungulados cinegéticos, el ciervo ibérico (*Cervus elaphus hispanicus*) y el gamo (*Dama dama*), en las fincas Lugar Nuevo (LN) y Selladores-Contadero (SC) en el sur de España. Para ello, utilizamos estimas de abundancia anuales de datos de densidad poblacional, caza y clima.

**Palabras clave:** Denso-dependencia, filtro de Kalman, gestión cinegética, modelo de espacio de estados.

## ABSTRACT

The detection of influential factors in populations dynamics of ungulates is very important for a specific hunting management program. One of the most important influential factors in the population dynamics is the density-dependence (DD), which acts as a self-regulation mechanism that affects recruitment and population growth rates in overabundance conditions. Consequently, detecting the presence of DD is essential to an appropriate harvest program. In this work a statistical model based on the Kalman filter has been used to detect the presence of DD, and show how climate variables and hunting pressure act on population dynamics in two populations of hunting ungulates, which are the Iberian red deer (*Cervus elaphus hispanicus*) and the fallow deer (*Dama dama*), in two reserves called Lugar Nuevo (LN) and Selladores-Contadero (SC) in southern Spain. To do this, we use estimates of annual abundance of population density, hunting and climate data.

**Keywords:** Density-dependence, Kalman filter, herd management, state-space model

# *Aplicación de la Estadística en la Evaluación de Competencias Escritoras*

Antonio de la Hoz Ruiz, Juan Antonio Maldonado Jurado

*Departamento de Estadística e Investigación Operativa, Universidad de Granada, e-mail:  
antoniodelahozruiz@gmail.com*

## RESUMEN

El estudio de las competencias escritoras expresadas por estudiantes universitarios ha adquirido en los últimos años una creciente atención en todas las disciplinas académicas. La demanda de múltiples textos escritos en la universidad contribuye a un cambio en la percepción de la enseñanza a este nivel; de hecho, la escritura universitaria es una competencia indispensable para la reflexión, adquisición y recreación del conocimiento científico/profesional.

Este Trabajo Fin de Grado fundamenta el diseño, la implementación y evaluación de un programa de escritura científico, online en la Universidad que permite crear nuevos contextos de aprendizaje donde se toma conciencia de las competencias escritoras metasociocognitivas a través del proceso de construcción de ensayos argumentativos.

En este estudio se utiliza una muestra compuesta por 187 estudiantes ( $n = 187$ ) de primer curso del Grado de Educación Primaria de la Facultad de Ciencias de la Educación de la Universidad de Granada. Se analizan Cuestionarios sobre Competencias Metasociocognitivas de la Escritura con un diseño pre-test y post-test; de manera similar se analizan Escalas de Percepción de la Autoeficacia en la Escritura del Ensayo Científico.

Por lo tanto, la finalidad de esta investigación es analizar si la aplicación del programa mejora la expresión de las competencias escritoras de los estudiantes universitarios, validar una escala junto con un cuestionario y comprobar en qué medida otras variables como la edad y el género puede determinar los efectos del programa utilizado sobre el aprendizaje de la escritura del estudiante.

**Palabras clave:** análisis estadístico, análisis factorial, estudio de correlación, competencias escritoras y autoeficacia.

## ABSTRACT

The study of the writing skills developed by undergraduates has caught the attention of all academic disciplines in recent years. The demand for multiple texts written at university level contributes to a change in the perception of teaching at this level. In fact, university writing is an indispensable skill for the reflection, acquisition and recreation of scientific/professional knowledge.

This end-of-degree paper is the basis for the design, implementation and evaluation of an online scientific writing programme at the University that allows the creation of new learning contexts where awareness of meta-socio cognitive writing skills is raised through the process of writing argumentative essays.

For this study, a sample of 187 students ( $n=187$ ) from the first year of Primary Education Studies of the Department of Educational Studies of the University of Granada is used.

Questionnaires on Meta-socio cognitive Writing Competences are analysed with a pre-test and a post-test design; similarly, Self-Efficacy Perception Scales in Scientific Essay Writing are analysed as well.

Therefore, the purpose of this research is to analyse whether the implementation of the programme improves the expression of the writing skills of university students, to validate a scale together with a specific questionnaire and to check to what extent other variables such as age and gender can determine the effects of the programme used on the students' learning of writing.

**Keywords:** statistical analysis, factor analysis, correlation studies, writing processes and self-efficacy.

# *Calidad de vida en personas mayores con síndrome de apnea del sueño. Relación con el deterioro cognitivo, deterioro funcional y enfermedades crónicas*

Lucía Ortega Donaire<sup>(1)</sup>, Sergio Iglesias Parro<sup>(2)</sup>, Ana Raquel Ortega Martínez<sup>(3)</sup>, María José Calero García<sup>(4)</sup>

(1) Departamento de Enfermería, Universidad de Jaén, e-mail: lortega@ujaen.es

(2) Departamento de Psicología, Universidad de Jaén, e-mail: siglesia@ujaen.es

(3) Departamento de Psicología, Universidad de Jaén, e-mail: arortega@ujaen.es

(4) Departamento de Enfermería, Universidad de Jaén, e-mail: mjcalero@ujaen.es

## RESUMEN

La apnea obstructiva del sueño (AOS) presenta una prevalencia del 30-80% en personas mayores de 65 años; a pesar de este hecho, es un síndrome que se encuentra infradiagnosticado y subtratado. La hipoxia y la fragmentación del sueño provocada por la AOS, puede ocasionar numerosas complicaciones en el funcionamiento cognitivo, así como en la calidad de vida de las personas que sufren esta enfermedad.

El propósito del presente estudio es determinar las posibles relaciones entre AOS y deterioro cognitivo y/o funcional, y la calidad de vida de la población adulta mayor, así como conocer si dichas relaciones están mediadas por variables como el grado de severidad de la AOS y la comorbilidad de enfermedades crónicas.

Se han recabado datos de 83 pacientes en la provincia de Jaén, los cuales acudían a la Unidad del Sueño del Hospital Doctor Sagaz. Los datos de dichos pacientes fueron recogidos a partir de instrumentos como: historia clínica (variables sociodemográficas), polisomnografía convencional e índice de apneas-hipopneas, Trail Making Test (TMT), Miniexamen cognoscitivo de Lobo, escala de Barthel para las actividades de la vida diaria y WHOQOL-BREF para la calidad de vida. Este es un estudio de casos y controles, con un análisis inferencial precedido de un análisis descriptivo de las puntuaciones obtenidas por cada participante en las diferentes variables. Se ha utilizado el coeficiente de correlación de Pearson para ver el grado de relación entre dos variables cuantitativas continuas. Los resultados obtenidos indican que los participantes con apnea del sueño en grado grave de nuestro estudio (20.5%), presentan relación con sufrir diabetes mellitus tipo II con un coeficiente de correlación (rp) de 0.25 ( $p < 0.05$ ), obesidad (rp=0.22,  $p < 0.05$ ), un alto IMC (rp=0.28,  $p = 0.01$ ), mayor número de centímetros de cintura (rp=0.26,  $p < 0.05$ ), mayor cansancio (rp=0.28,  $p < 0.01$ ), somnolencia diurna (rp=0.33,  $p < 0.01$ ), menor puntuación en el dominio "energía" de la escala WHOQOL-BREF que evalúa la calidad de vida (rp=-0.32,  $p < 0.01$ ) y escala Barthel (rp=-0.23,  $p < 0.05$ ). Por otra parte, se obtiene una puntuación TMT-A rp=-0.02,  $p = 0.877$ ; TMT-B rp=0.087,  $p = 0.48$  y en el miniexamen de Lobo rp=-0.05,  $p = 0.673$ .

En conclusión, padecer apnea del sueño en personas mayores de 65 años está asociado a cambios en su bienestar físico, mental y social, especialmente las personas que padecen AOS en grado grave, ya que son las que más complicaciones acarrear. La prevención,

educación para la salud y un diagnóstico precoz, son elementos clave en este síndrome y grupo poblacional.

**Palabras clave:** Persona de edad avanzada; apnea obstructiva del sueño; deterioro cognitivo; deterioro funcional; calidad de vida; comorbilidades.

## ABSTRACT

Obstructive sleep apnoea (OSA) has a prevalence of 30-80% in people over 65 years of age; despite this fact, it is an under-diagnosed and undertreated syndrome. Hypoxia and sleep fragmentation caused by OSA can cause numerous complications in cognitive functioning as well as in the quality of life of people suffering from this disease.

The purpose of the present study is to determine the possible relationships between OSA and cognitive and/or functional impairment, and the quality of life of the older adult population, as well as to know if these relationships are mediated by variables such as the degree of severity of OSA and the comorbidity of chronic diseases.

Data have been collected from 83 patients in the province of Jaen, who came to the Sleep Unit of the Hospital Doctor Sagaz. The data of these patients were collected from instruments such as: clinical history (sociodemographic variables), conventional polysomnography and apneas-hypopneas index, Trail Making Test (TMT), Lobo's cognitive mini-examination, Barthel's scale for activities of daily life and WHOQOL-BREF for quality of life. It is a study of cases and controls, with an inferential analysis preceded by a descriptive analysis of the scores obtained by each participant in the different variables. Pearson's correlation coefficient has been used to see the degree of relationship between two continuous quantitative variables. The results obtained indicate that the participants with severe sleep apnea in our study (20.5%) have a relationship with type II diabetes mellitus with a correlation coefficient ( $r_p$ ) of 0.25 ( $p < 0.05$ ), obesity ( $r_p = 0.22$ ,  $p < 0.05$ ), a high BMI ( $r_p = 0.28$ ,  $p = 0.01$ ), higher number of waist centimeters ( $r_p = 0.26$ ,  $p < 0.05$ ), higher fatigue ( $r_p = 0.28$ ,  $p < 0.01$ ), daytime drowsiness ( $r_p = 0.33$ ,  $p < 0.01$ ), lower score in the "energy" domain of the WHOQOL-BREF scale assessing quality of life ( $r_p = -0.32$ ,  $p < 0.01$ ) and Barthel scale ( $r_p = -0.23$ ,  $p < 0.05$ ). On the other hand, a score is obtained TMT-A  $r_p = -0.02$ ,  $p = 0.877$ ; TMT-B  $r_p = 0.087$ ,  $p = 0.48$  and in Lobo's miniexamination  $r_p = -0.05$ ,  $p = 0.673$ . In conclusion, suffering from sleep apnea in people over 65 is associated with changes in their physical, mental and social well-being, especially people who suffer from severe OSA, as they are the most complications. Prevention, health education and early diagnosis are key elements in this syndrome and population group.

**Keywords:** aged; sleep, apnea obstructive; cognitive dysfunction; frail elderly; quality of life; comorbidities.

# ***Encefalopatía Epiléptica STXBP1: predictores computacionales y estadísticos de asociación genotipo/fenotipo***

Isabel Trigo Pérez<sup>(1)</sup>, Francisco J. Esteban<sup>(2)</sup>

(1) Dep. Biología Experimental, Universidad de Jaén, e-mail: itp00005@red.ujaen.es

(2) Dep. Biología Experimental, Universidad de Jaén, e-mail: festeban@ujaen.es

## **RESUMEN**

Las encefalopatías epilépticas infantiles de inicio temprano (EIEE, Orphan #1934) se caracterizan por crisis médicamente refractarias, alteraciones graves de la actividad cerebral, encefalopatía difusa y retraso del desarrollo psicomotor severo. Entre ellas, la EIEE tipo 4 (OMIM #612164), se debe a mutaciones en el gen STXBP1 (también denominado MUNC18-1). Los pacientes con síndrome STXBP1 suelen, además, presentar alteraciones en el sueño, espasmos, mioclonía, ataxia, disfunciones cardíacas, trastornos de espectro autista y esquizofrenia. Su diagnóstico está pobremente desarrollado y los diferentes tipos de mutaciones encontradas (genotipo) aún no están bien correlacionadas con las manifestaciones clínicas de cada individuo (fenotipo). En este sentido, pacientes con una misma mutación pueden mostrar diferentes síntomas y comorbilidades, e individuos con una sintomatología similar pueden ser portadores de un tipo diferente de mutación. A ello hay que añadir la diferente respuesta que presentan al tratamiento. Todo esto conlleva la falta de un correcto asesoramiento genético y asistencia a los padres, su alarma y un retraso sustancial en la correcta aplicación de tratamientos adecuados, lo que repercute gravemente en el pronóstico de estos niños.

Para contribuir a solucionar esta situación, estamos llevando a cabo un análisis computacional de relaciones genotipo/fenotipo basado en la Biomedicina de Sistemas, es decir, mediante la integración y el análisis de datos masivos, clínicos y moleculares, a distintos niveles utilizando principalmente métodos computacionales. Para ello, accedemos a la información biomédica y genómica pública disponible, como por ejemplo la incorporada en la plataforma ClinVar de la NBCI ([ncbi.nlm.nih.gov/clinvar](http://ncbi.nlm.nih.gov/clinvar)) sobre variantes clínicas, y el catálogo online de genes humanos ExAC ([exac.broadinstitute.org](http://exac.broadinstitute.org)), el cual incluye a 60,706 individuos secuenciados como parte de diferentes estudios tanto en enfermedades específicas como genéticos poblacionales.

Los datos relacionales complejos considerados de interés se extraen mediante el uso de scripts personalizados programados en los lenguajes R ([r-project.org](http://r-project.org)) y Python ([python.org](http://python.org)). En este sentido, y con el fin de conocer con todo detalle el número y relación fenotípica de las variantes genéticas asociadas a STXBP1, se obtienen todas las entradas para este gen en cada una de las bases de datos anteriores y, de cada una de ellas, con todas sus variables asociadas y las categorías descriptivas. A partir de estas entradas podemos, por ejemplo, conocer el número de variantes por grado de significación clínica, los tipos de mutación (*missense*, *nonsense*, *indel*, etc.), los fenotipos asociados, la localización cromosómica, etc., y con ellas construimos una base de datos única para STXBP1 teniendo en cuenta todas las variables de interés de todas las bases de datos, compilando información no coincidente y eliminando contenidos redundantes.

Con la información contenida en esta base de datos unificada desarrollamos los predictores computacionales de asociación genotipo/fenotipo. Se generan clasificadores y predictores computacionales para tratar de detectar y establecer las posibles relaciones existentes genotipo/fenotipo, y las comorbilidades asociadas a cada tipo de mutación, mediante el uso de algoritmos tales como SVM, KNN, modelos logísticos, redes neuronales, árboles de clasificación y redes bayesianas. De modo previo, tanto las mutaciones como los fenotipos se priorizan mediante selección de atributos por información mutua, entropía y ganancia de información, y los análisis se llevan a cabo con el uso de WEKA ([cs.waikato.ac.nz/ml/weka](http://cs.waikato.ac.nz/ml/weka)) y Python ([python.org](http://python.org)), en este último caso mediante la generación de scripts personalizados con funciones ya implementadas en paquetes específicos (Numpy, Pandas, Matplotlib, Seaborn, y Scikit-Learn). La idoneidad de clasificación y predicción se establece a través de la información proporcionada por matrices de confusión, área bajo curvas ROC y estadístico kappa, entre otros, previa aplicación de entrenamientos y clasificaciones tipo *10-fold cross-validation* o *leave-one-out* según el tipo y cantidad de datos.

**Palabras clave:** Análisis de datos masivos, clasificadores, minería de datos, Munc18-1, predictores, syntaxin binding protein 1.

## ABSTRACT

Early-onset childhood epileptic encephalopathies (EIEE, Orphan # 1934) are characterized by medically refractory seizures, severe alterations in brain activity, diffuse encephalopathy and delayed severe psychomotor development. Among them, EIEE type 4 (OMIM # 612164) is due to mutations in the STXBP1 gene (also called MUNC18-1). Patients with STXBP1 syndrome also suffer sleep disorders, spasms, myoclonus, ataxia, cardiac dysfunctions, autism spectrum disorders and schizophrenia. The diagnosis is poorly developed, and the different types of mutations found (genotype) are not yet well correlated with the clinical manifestations of each individual (phenotype). In this sense, patients with the same mutation may show different symptoms and comorbidities; and individuals with similar symptoms may be carriers of a different type of mutation. Moreover, patients show a different response to treatments. Taken together, an inadequate genetic counseling and assistance to parents is produced, with the corresponding alarm and a delay in the appropriated treatments, which has a serious impact on the prognosis of these children.

With the final aim to help to unravel this situation, we are carrying out a computational analysis of genotype/phenotype relationships based on Systems Biomedicine, that is, through the integration and analysis of massive, clinical and molecular data, at different levels mainly using computational methods. For it, we access to the public biomedical and genomic information available, such as the one incorporated in the NCBI ClinVar platform ([ncbi.nlm.nih.gov/clinvar](http://ncbi.nlm.nih.gov/clinvar)), about clinical variants, and the online catalog of ExAC human genes ([exac.broadinstitute.org](http://exac.broadinstitute.org)), which includes 60,706 individuals sequenced as part of different studies in both specific diseases and population genetics.

Specific complex relational data of interest can be obtained using custom scripts in the R ([r-project.org](http://r-project.org)) and Python ([python.org](http://python.org)) languages. In this sense, and in order to know in detail the number and phenotypic relationship of the genetic variants associated with STXBP1, all entries for this gene in each of the previous databases are obtained including all the associated variables and descriptive categories. From these entries we can get, for example, the number of variants by grade of clinical significance, the types of mutation (nonsense, meaningless, indel, etc.), the associated phenotypes, the chromosomal location, etc., and taken together, to build a unique database for STXBP1 including all the variables of interest of all databases, compiling mismatched information and eliminating redundant contents.

With the information contained in this unified database, we develop the computational classifiers and predictors of genotype/phenotype associations, trying to detect not only the possible genotype/phenotype affected relationships, but also the comorbidities associated with each type of mutation. To do this, algorithms such as SVM, KNN, logistic models, neural networks, classification trees and Bayesian networks are chosen as the appropriate tools. Previously, both mutations and phenotypes are prioritized through the selection of attributes by mutual information, entropy and information gain, using WEKA ([cs.waikato.ac.nz/ml/weka](http://cs.waikato.ac.nz/ml/weka)) and Python ([python.org](http://python.org)), generating custom scripts with functions implemented in specific packages (Numpy, Pandas, Matplotlib, Seaborn, and Scikit-Learn). Classification and prediction suitability is obtained from the information specified by confusion matrices, area under ROC curves and kappa statistics, among others, with the application of training and testing with 10-fold cross-validation or leave-one-out procedures, according to the type and amount of data.

**Keywords:** Massive data analysis, classifiers, data mining, Munc18-1, predictors, syntaxin binding protein 1.

# *Estudio de la validez de la Autoevaluación como método de evaluación educativa: uso del bootstrapping*

Samuel P. León<sup>(1)</sup>

(1) Departamento de Pedagogía, Universidad de Jaén, [sparra@ujaen.es](mailto:sparra@ujaen.es):

## RESUMEN

La autoevaluación es una práctica muy extendida tanto en el ámbito de la investigación como en la práctica docente. No obstante, hay muchas incógnitas alrededor de este tipo de evaluación académica. El objetivo principal del presente trabajo es examinar empíricamente la validez de la autoevaluación como prueba de evaluación académica en estudiantes universitarios. Para estudiar dicha validez, los datos reportados por los estudiantes por medio de la autoevaluación se contrastaron con los datos propuestos por la evaluación del profesor experto (comúnmente denominado *gold standard*). En primer lugar, los resultados fueron analizados mediante regresión lineal. Dado que la muestra fue un grupo natural de 64, se propone un análisis basado en bootstrapping para estimar el efecto poblacional. Los resultados encontrados nos permitirán concluir que la autoevaluación, en los términos propuestos, puede ofrecer una validez moderada como prueba de evaluación académica en términos generales, pero esta validez decaerá cuando el análisis se realiza agrupando a los alumnos por su rendimiento académico. En este caso, cuando analizamos los resultados agrupando a los alumnos por su rendimiento en la prueba, los resultados revelaron que la autoevaluación mostraba un sesgo hacia la sobrevaloración en los alumnos menos capaces, mientras que en el caso de los alumnos más capaces este sesgo tornaba en infravaloración. Las futuras investigaciones deberían de centrarse en profundizar sobre los factores que pueden atentar a la validez de la autoevaluación dentro del contexto académico.

**Palabras clave:** Autoevaluación, validez, educación universitaria, bootstrap

## ABSTRACT

Self-assessment is a widespread practice both in the field of research and in teaching practice. However, there are many unknowns around this type of academic evaluation. The main objective of this paper is to empirically examine the validity of the self-assessment as proof of academic evaluation in university students. To study this validity, the data reported by the students by means of the self-assessment were contrasted with the data proposed by the evaluation of the expert teacher (commonly called *gold standard*). First, the results were analysed by linear regression. Since the sample was a natural group of 64, an analysis based on bootstrapping is proposed to estimate the population effect. The results found will allow us to conclude that the self-assessment, in the proposed terms, may offer moderate validity as a test of academic evaluation in general terms, but this validity will decline when the analysis is carried out by grouping the students by their academic performance. In this case, when we analysed the results by grouping the students by their performance on the test, the results revealed that the self-assessment showed a bias towards overvaluation in the less

able students, while in the case of the most capable students this bias became undervaluation. Future research should focus on deepening the factors that may undermine the validity of the self-assessment within the academic context.

**Keywords:** Self-assessment, validity, university education, bootstrap

# ***Regresión Logística Multinivel Multinomial aplicada a estudios en turismo***

Julio Vena Oya<sup>(1)</sup>, José Alberto Castañeda García<sup>(2)</sup>, Miguel Ángel Rodríguez Molina<sup>(3)</sup>

(1) Departamento de Comercialización e Investigación de Mercados, Universidad de Granada, e-mail: juliovena@correo.ugr.es

(2) Departamento de Comercialización e Investigación de Mercados, Universidad de Granada, e-mail: jalberto@ugr.es

(3) Departamento Comercialización e Investigación de Mercados, Universidad de Granada, e-mail: rmolina@ugr.es

## **INTRODUCCIÓN**

A la hora de analizar los datos debemos pensar en la estructura de los mismos. Un caso de estudio recurrente es aquel en el que cada turista realiza una serie de gastos durante su estancia, lo que nos lleva a pensar que nuestros datos presentan una estructura jerárquica o anidada. Como es lógico, todos los gastos de un mismo turista estarán relacionados con las características de éste. Puesto que cada conjunto de gastos viene dado por un mismo turista, se hace imposible mantener la independencia de las observaciones ya que, al pertenecer a un mismo individuo, los gastos estarán correlacionados. Para analizar este tipo de estructuras anidadas de datos la herramienta adecuada es la regresión multinivel (Hox, 2013). Una variante de los modelos de regresión multinivel es aquella que permite incluir como dependiente variables dicotómicas o categóricas. Tradicionalmente, para respuestas de tipo categórico, se han venido usando métodos de MCO, introduciendo alguna modificación en la variable. Sin embargo, es cada vez más frecuente usar modelos especialmente diseñados para este tipo de variables, como son los modelos de regresión logística para respuestas binomiales o modelos logit generalizados para respuestas multinomiales (Agresti et al., 2000).

## **OBJETIVO**

El objetivo de este estudio es identificar los determinantes del gasto turística durante la visita a Granada, con un enfoque multinivel. EL software elegido para el análisis de los datos fue IBM SPSS Statistics, versión 21.

Para ello seguiremos el proceso de modelaje descrito por Heck et al. (2013, págs. 261-327), los cuales dedican un capítulo para este tipo de modelos empleando dicho software.

## **RESULTADOS**

Se procede pues a la estimación de los diferentes modelos descritos por Heck et al. (2013) dando lugar al siguiente modelo de regresión logística multinivel multinomial, con determinantes del gasto a nivel del turista y a nivel del propio gasto:

Tabla 1: Coeficientes estimados (significación unilateral)

	Gasto bajo	Gasto mediano	Gasto superior
Intersección	-1.7 (0.52)	-3.05 (0.00)	-3.4 (0.00)
Edad	0.08 (0.00)	0.05 (0.04)	0.08 (0.00)
Motivo de la visita	0.06 (0.19)	0.24 (0.00)	0.18 (0.00)
Nº visitas previas	-0.41 (0.01)	-0.348 (0.03)	-0.20 (0.13)
Momento del gasto	-0.35 (0.04)	-0.47 (0.01)	-0.48 (0.02)
Tipo de gasto	3.81 (0.00)	4.99 (0.00)	4.52 (0.00)
Distancia entre gastos (Km)	0.44 (0.07)	0.60 (0.02)	0.52 (0.05)

El procedimiento de selección del modelo óptimo será explicado con mayor profundidad en las sesiones paralelas.

Estos resultados no pueden ser interpretados directamente, si no que requieren de una transformación exponencial (Aldás y Uriel, 2005), dando como resultado la razón de apuesta u odds ratio, el cual indica la ventaja (o desventaja) que otorga una variable para realizar un gasto en ese tramo, frente a la categoría de referencia (no gastar).

## BIBLIOGRAFÍA

Agresti, A. (2000). *Random-effects modeling of categorical response data*. Sociological Methodology, 30(1), 27-80.

Aldás, J., & Uriel, E. (2005). *Análisis multivariante aplicado. Aplicaciones Al Marketing, Investigación De Mercados, Economía, Dirección De Empresas y Turismo*. Ed.Thomson,

Heck, R. H., Thomas, S., & Tabata, L. (2013). *Multilevel modeling of categorical outcomes using IBM SPSS*. Routledge Academic.

Hox, J. J. (2013). *Multilevel regression and multilevel structural equation modeling*. The Oxford Handbook of Quantitative Methods, 2(1), 281-294.

**Palabras clave:** Regresión logística multinivel multinomial, turismo, gasto turístico.

# *Alfabetización estadística y fake news, una herramienta contra la desinformación*

Molina-Portillo, Elena<sup>(1)</sup>, Martínez, Francisco<sup>(2)</sup>, Bizet, Valeria<sup>(3)</sup>, Danilo Díaz Levicoy<sup>(4)</sup>,  
y Contreras, José Miguel<sup>(5)</sup>

(1) Departamento de Didáctica de la Matemática, Universidad de Granada, elemo@ugr.es

(2) Departamento de Didáctica de las Ciencias, Universidad de Jaén,

(3) Estudiante de doctorado, Universidad de Granada, valeriabizet@gmail.com

(4) Departamento de Matemática, Física y Estadística, Universidad Católica del Maule,  
dddiaz01@hotmail.com

(5) Departamento de Didáctica de la Matemática, Universidad de Granada, jmcontreras@ugr.es

## RESUMEN

Hoy en día las Fakes News, técnicamente “desinformación”, son una autentica plaga que afectan tanto a las decisiones personales como a colectivas. Su repercusión, debido a la proliferación de las redes sociales, ha sido considerado como un ejemplo de la vulnerabilidad y debilidad democrática ante un aumento de la manipulación intencional y sistemática de la información.

Las noticias falsas se emiten con la intención deliberada de manipular, engañar o inducir a errores en las decisiones personales, desprestigiando instituciones o personas con la para sacar algún tipo de rédito de ello. Es por ello que, en los últimos años, ha habido una concienciación respecto este tema por parte de políticos, expertos y periodistas, que han tratado de proponer iniciativas para contrarrestar el efecto de las fake news en la sociedad. Estas iniciativas van desde la denuncia de su utilización y efecto, intentando crear un debate social sobre el tema en cuestión, a acciones estratégicas promovidas por organismos internacionales. Por ejemplo, la Unión Europea constituyó en 2018 un grupo de expertos para tratar el tema con mayor profundidad. En su informe se conciben algunas acciones destinadas a mejorar la transparencia de la información, que van desde el desarrollo de herramientas para abordar la desinformación, a iniciativas destinadas a promover la alfabetización mediática y la promoción de la investigación sobre el impacto de la desinformación en la sociedad.

Hay un acuerdo general en el que, para tratar la alfabetización mediática, hay que tener en cuenta otros tipos de alfabetizaciones más allá de la periodística, tales como la alfabetización informacional, cívica, informal, funcional, digital, tecnológica, etc., ya que ocupan un terreno conceptual compartido en el fenómeno de las fake news. En nuestro caso, nos centramos en la alfabetización estadística, marco muy ligado a la alfabetización mediática ya que la estadística en general, y los datos en particular, fundamentan y dan validez a la información. Los datos legitiman información de investigaciones, estudios, encuestas, etc., que la sociedad asume por ciertos o exactos simplemente porque esta información está validada por las estadísticas (Hofstadter, 2008). Pero como ya incidió Bauman (2005), somos muchas veces víctimas del abuso de los datos, debido a que aceptamos su uso, y muchas veces su abuso, para explicar cualquier fundamentación aprovechando la inalterable objetividad de ellos. Lo que lleva, en muchas ocasiones al uso de datos para fundamentar noticias falsas y que estas sean más creíbles.

En este trabajo, además de mostrar ejemplos de la problemática, se pretende mostrar el grado de alfabetización estadística que poseen un grupo de estudiantes 1º de bachillerato en función de cómo sea presentada un tipo de información sesgada, presentada en distintas representaciones, en

concreto: frecuencias, representaciones gráficas, porcentajes y probabilidad, y cómo influyen estas representaciones en la interpretación que se realice de estas noticias manipuladas.

**Palabras clave:** Desinformación, Fake News, Alfabetización estadística, Educación secundaria.

## ABSTRACT

Today the Fakes News, technically "misinformation", are a real plague that affect both personal and collective decisions. Its impact, due to the proliferation of social networks, has been considered as an example of democratic vulnerability and weakness in the face of an increase in the intentional and systematic manipulation of information.

Fake news is broadcast with the deliberate intention of manipulating, deceiving or inducing errors in personal decisions, discrediting institutions or people with the to get some kind of revenue from it. That is why, in recent years, there has been an awareness of this issue by politicians, experts and journalists, who have tried to propose initiatives to counteract the effect of fake news on society. These initiatives range from the denunciation of their use and effect, trying to create a social debate on the subject in question, to strategic actions promoted by international organizations. For example, the European Union formed a group of experts in 2018 to address the issue in greater depth. In its report some actions are designed to improve the transparency of information, ranging from the development of tools to address disinformation, to initiatives aimed at promoting media literacy and promoting research on the impact of disinformation on society.

There is a general agreement in which, in order to deal with media literacy, other types of literacy beyond journalism must be considered, such as informational, civic, informal, functional, digital, technological literacy, etc., since they occupy a shared conceptual terrain in the phenomenon of fake news. In our case, we focus on statistical literacy, a framework closely linked to media literacy since statistics in general, and data in particular, support and validate the information. The data legitimizes information from research, studies, surveys, etc., that society assumes for certain or exact reasons simply because this information is validated by statistics (Hofstadter, 2008). But as Bauman (2005) already affected, we are often victims of data abuse, because we accept their use, and often their abuse, to explain any foundation taking advantage of their unalterable objectivity. What leads, in many occasions to the use of data to base false news and that these are more credible.

In this work, in addition to showing examples of the problem, it is intended to show the degree of statistical literacy that a group of high school students possess according to how a type of biased information is presented, presented in different representations, specifically: frequencies, graphical representations, percentages and probability, and how these representations influence the interpretation that is made of this manipulated news.

**Keywords:** misinformation, Fake News, Statistical Literacy, Secondary Education.

# *Análisis de alteraciones neurológicas complejas mediante teoría de juegos cooperativos aplicada a datos de microarrays*

José A. Castro<sup>(1)</sup>, Francisco J. Esteban<sup>(2)</sup>

(1) Departamento de Biología Experimental, Universidad de Jaén, e-mail: jacm0013@red.ujaen.es

(2) Departamento de Biología Experimental, Universidad de Jaén, e-mail: festeban@ujaen.es

## RESUMEN

La tecnología de microarrays permite un análisis completo de la expresión génica, y es ampliamente utilizado en investigación biomédica para detectar expresión génica alterada de genes particulares en una enfermedad dada en comparación con controles sanos. El análisis del componente genético de alteraciones neurológicas complejas como el autismo, la esquizofrenia, la depresión y el trastorno bipolar, resulta particularmente complicado y muestra una falta general de concordancia en los resultados obtenidos a partir de experimentos de microarrays. Los enfoques en el procesamiento de datos de microarrays se basan generalmente en un ordenamiento de genes de acuerdo a sus p-valores, pero p-valores pequeños no son necesariamente indicadores de una señal biológica, y p-valores elevados, supuestamente insignificantes, pueden estar relacionados con genes implicados en procesos biológicos clave. Además, el procesamiento estadístico excesivo de los p-valores, que a menudo se aplica para corregir la detección de falsos positivos y la pérdida de señal, tiende a ser demasiado conservadora para detectar resultados informativos biológicamente.

Motivados por la necesidad de enfoques más poderosos en la detección de señal, un método para el análisis de expresión génica basado en la teoría de juegos fue propuesto por Moretti y colaboradores (Moretti et al. 2007, TOP 15: 265-280). Este enfoque está basado en la computación del valor de Shapley, un índice numérico que representa la relevancia de cada gen en una condición concreta mientras que se tiene en cuenta simultáneamente el comportamiento de expresión de todos los demás genes evaluados en las mismas condiciones. A partir de este método, en combinación con la estadística, se ha demostrado que resulta de utilidad para el análisis de los datos de expresión génica diferenciales. En nuestro grupo de investigación se ha aplicado esta metodología para determinar la relevancia génica de un experimento de microarrays en autismo, en el cual sólo unos pocos genes podían ser detectados utilizando métodos estadísticos convencionales (Esteban and Wall. 2009, TOP 19: 121-129).

En este trabajo presentamos los resultados obtenidos en el análisis de diferentes datasets (GSE6575, GSE18123, GSE25507, GSE17612, GSE62333, GSE5389, GSE7036, GSE12654 y GSE53987), descargados de Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>), correspondientes a estudios de autismo, esquizofrenia, trastorno bipolar, depresión mayor y población control. Los análisis estadísticos han sido realizados con Bioconductor (<http://www.bioconductor.org/>) en R (<http://www.r-project.org/>). Para obtener la relevancia de los genes en un conjunto de datos de microarrays utilizamos un método de juegos de coalición como se describió

previamente (Esteban and Wall. 2009, TOP 19: 121-129). El enriquecimiento de las funciones biológicas y la sobrerrepresentación de las enfermedades dentro de los términos génicos se detectaron utilizando g:profiler (<https://biit.cs.ut.ee/gprofiler/gost>).

Los resultados obtenidos mostraron la detección de un bajo número de genes en cada experimento, con la única excepción de GSE18123. Por otra parte, fue posible obtener genes con un p-valor menor a 0.01 y 0.05, pero la mayor parte de éstos fueron rechazados debido a la corrección de Bonferroni. Otros enfoques como FDR (False Discovery Rate) podían aumentar el número de genes detectados significativamente, pero asumiendo una proporción significativa de falsos positivos. Por otro lado, la técnica de juegos de microarrays nos permite obtener grandes cantidades de sondas candidatas en cada análisis y para cada una de los cuatro fenotipos de interés, incluso más de 2000 sondas en un solo experimento, pero hay una gran variación en el número de sondas significativas, dependiendo del experimento y el fenotipo estudiado. El análisis de enriquecimiento funcional ha detectado la alteración de procesos biológicos tales como funciones neuronales, endocrinas e inmunes, así como aquellos relacionados con metabolismo energético a nivel celular, y procesos de señalización celular.

En conclusión, el comportamiento de expresión de los genes detectados, aplicando teoría de juegos cooperativos, es una buena aproximación para obtener conocimiento biológico en la compleja base molecular los trastornos neurológicos complejos.

**Palabras clave:** juegos de coalición, trastornos complejos, valor de Shapley, juegos de microarrays

## ABSTRACT

Microarrays technology allows a complete genome-wide analysis of the gene expression, and it is widely used in biomedical research to detect altered gene expression of particular genes in a given disease when compared to healthy controls. The genetic component of complex neurological multigenic disorders such as autism, schizophrenia, bipolar disorder and major depression, show a general absence of concordance in the obtained results. Microarray data processing approaches are generally based on gene ranking according to p-values, but small p-values are not necessarily indicative of biological signal, and large, presumably insignificant p-values, may be related to genes playing key roles. In addition, the further statistical processing of p-values, which is often applied to correct for false discovery and loss of signal, use to be too conservative to detect biologically informative results.

Motivated by the need for more powerful approaches to signal detection, a method for gene expression analysis based on game theory was proposed by Moretti and co-workers (Moretti et al. 2007, TOP 15: 265-280). This approach is based on the computation of the Shapley value, a numerical index which represents the relevance of each gene under a certain condition while simultaneous accounting for the expression behaviors of the other genes under the same condition. Because this method, in combination with statistics, has been demonstrated to be useful for differential gene expression data analysis, we applied this methodology to determine the gene relevance on an ASD microarray experiment in

which only a few genes could be detected using standard statistical approaches (Esteban and Wall. 2009, TOP 19: 121-129).

Briefly, we have analyzed several subsets from experiments GSE6575, GSE18123, GSE25507, GSE17612, GSE62333, GSE5389, GSE7036, GSE12654 and GSE53987, downloaded from Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>). These subsets consisted in several experiments that harbor samples obtained from patients of autism, schizophrenia, bipolar disorder, major depression and control population. Statistical analyses were performed with Bioconductor (<http://www.bioconductor.org/>) in R (<http://www.r-project.org/>). To obtain the gene relevance on the microarray data-sets, we used the coalitional game method as previously described (Esteban and Wall. 2009, TOP 19: 121-129). Enrichment of biological function(s) and overrepresentation of disease(s) among gene terms were detected using g:profiler (<https://biit.cs.ut.ee/gprofiler/gost>).

The obtained results showed a low number of genes differentially expressed in each experiment, with the only exception of GSE18123. On the other hand, it was possible to obtain genes with a raw p-value  $< 0.01$  and  $0.05$ , but most of them were rejected after Bonferroni correction. Other approaches as FDR (False Discovery Rate) can increase the number of significant genes, but increasing the proportion of false positives. On the other hand, the microarray game did increase the candidate probes in each analysis and phenotypes, even more that 2000 probes in only one experiment, but there is a great variation in the number of significative probes, depending of the experiment and studied phenotype. Gene Ontology analysis has detected neuronal, endocrine and immune functions, also those related to metabolism at the cellular level, and cellular signaling functions, among other several biological pathways.

In conclusion, the expression behavior of the genes detected with a microarray game is a useful approach to obtain biological knowledge in the complex molecular basis of complex neurological disorders.

**Keywords:** coalitional games, complex disorders, Shapley value, microarray game.

# *La microestructura en los libros de la enseñanza secundaria obligatoria: las medidas de dispersión*

Jesús Del Pino Ruiz <sup>(1)</sup>, Antonio Estepa Castro <sup>(2)</sup>

(1) *Didáctica de las Ciencias, Universidad de Jaén, jpr00026@red.ujaen.es:*

(2) *Didáctica de las Ciencias, Universidad de Jaén, aestepa@ujaen.es:*

## RESUMEN

Este trabajo es parte de otro más amplio referido al estudio de las medidas de dispersión en 3º y 4º cursos de la Educación Secundaria Obligatoria (ESO) en España. Comienza con una breve exposición de la investigación de la estructura en los libros de texto de Matemáticas, ejemplificado brevemente con libros de Reino Unido, Francia y Alemania. Se continúa con una exposición de las medidas de dispersión en el currículo actual. Se describe la metodología que es exploratoria. Se estudia la estructura de los libros de la muestra, describiendo brevemente la macroestructura y realizando un estudio más detallado de la microestructura y los contenidos de las medidas de dispersión en los libros de la muestra. Finalmente, se obtiene unas conclusiones de interés, creemos, para la enseñanza del tema.

**Palabras clave:** libros de texto, estructura de los libros de texto, macroestructura, microestructura, medidas de dispersión.

## ABSTRACT

This work is part of a broader one referring to the study of dispersion measures in 3rd and 4th year of Compulsory Secondary Education (ESO) in Spain. It begins with a brief exposition of the structure research in Mathematics textbooks, briefly exemplified by books from the United Kingdom, France and Germany. It continues with an exhibition of the measures of dispersion in the current curriculum. The methodology that is exploratory is described. The structure of the sample books is studied, briefly describing the macro structure and carrying out a more detailed study of the microstructure and the contents of the dispersion measures in the sample books. Finally, we get some conclusions of interest, we believe, for teaching the subject.

**Keywords:** textbooks, structure of textbooks, macro structure, microstructure, dispersion measures.

# *Indicadores de transparencia para organizaciones sin ánimo de lucro: Estudio de un caso español*

Cristina Ortega-Rodríguez<sup>(1)</sup>, Antonio Luis Moreno Albarracín<sup>(2)</sup>, Ana Licerán-Gutierrez<sup>(3)</sup>, Álvaro Labella Romero<sup>(4)</sup>

(1) Departamento de Economía Financiera y Contabilidad: [corodrig@ujaen.es](mailto:corodrig@ujaen.es)

(2) Departamento de Economía Financiera y Contabilidad: [almoreno@ujaen.es](mailto:almoreno@ujaen.es)

(3) Departamento de Economía Financiera y Contabilidad; [aliceran@ujaen.es](mailto:aliceran@ujaen.es)

(4) Departamento de Informática; [alabella@ujaen.es](mailto:alabella@ujaen.es)

## RESUMEN

**Introducción:** Las organizaciones sin ánimo de lucro (ESAL, en adelante) juegan un papel clave en nuestra sociedad debido a las actividades sociales que desarrollan, recibiendo para ello fondos tanto de entidades públicas como privadas. En estos términos, la transparencia de la información que divulgan las ESAL a sus grupos de interés sobre la gestión de los fondos que reciben, tiene un interés creciente. A pesar de la importancia de la transparencia y su amplio análisis en la literatura, el hecho es que pocas investigaciones han medido empíricamente el alcance de la transparencia en las ESAL.

**Objetivo:** Para llenar este vacío, nuestra investigación tiene como objetivo validar empíricamente la ponderación de los indicadores propuestos por la Coordinadora de ONG para el Desarrollo (CONGDE) y el documento de la Plataforma de Acción Social ONG con las opiniones de los expertos en el campo del Tercer Sector. Para este propósito, utilizamos el Método Best-Worst (BWM), una técnica novedosa, ampliamente utilizada y formal, para asignar pesos óptimos a los criterios en entornos de toma de decisiones con criterios múltiples. BWM (Rezaei, 2015) es un método de toma de decisiones grupales multicriterio (MCDM- Multi-criteria group decision making) que calcula la priorización de diferentes elementos de decisión a través de comparaciones por pares. Esta técnica consiste en comparar solo el mejor y el peor elemento del problema de decisión, ambos seleccionados por los tomadores de decisiones, con el resto de ellos.

**Resultados y discusión:** Nuestros resultados (Ver TABLA 1) muestran (en valores medios y divididos en 4 bloques) diferencias interesantes en los pesos de cada uno de los indicadores por parte de la CONGDE y los expertos de las ESAL, lo que sugiere una revisión de la importancia propuesta para que los indicadores midan el grado de transparencia.

TABLA 1. Valores medios según CONGDE y opinión de expertos.

		BLOQUE 1. ÓRGANO DE GOBIERNO									
		TR1.1	TR1.2	TR1.3	TR1.4	TR1.5					
VALOR MEDIO	BWM	0,169	0,118	0,193	0,373	0,146					
VALOR MEDIO	CONGDE	0,3	0,2	0,1	0,3	0,1					
		BLOQUE 2. MISIÓN, VISIÓN Y VALORES									
		TR2.1	TR2.2	TR2.3	TR2.4	TR2.5					
VALOR MEDIO	BWM	0,299	0,202	0,192	0,148	0,159					
VALOR MEDIO	CONGDE	0,3	0,2	0,25	0,1	0,15					
		BLOQUE 3. BASE SOCIAL Y APOYOS									
		TR3.1	TR3.2	TR3.3	TR3.4	TR3.5	TR3.6	TR3.7			
VALOR MEDIO	BWM	0,126	0,086	0,207	0,124	0,168	0,118	0,172			
VALOR MEDIO	CONGDE	0,25	0,15	0,15	0,1	0,15	0,100	0,1			
		BLOQUE 4. PLANIFICACIÓN Y RENDICIÓN DE CUENTAS									
		TR4.1	TR4.2	TR4.3	TR4.4	TR4.5	TR4.6	TR4.7	TR4.8	TR4.9	TR4.10
VALOR MEDIO	BWM	0,087	0,074	0,2	0,1	0,12	0,081	0,061	0,082	0,12	0,075
VALOR MEDIO	CONGDE	0,1	0,1	0,15	0,1	0,05	0,1	0,1	0,1	0,1	0,1

Fuente: Elaboración propia.

**Conclusión:** Nuestra contribución es doble: desde el punto de vista teórico, hemos detectado que los sistemas de contabilidad actuales en las ESAL no son suficientes hoy en día para tomar una decisión precisa y aceptable por parte de las partes interesadas, así como para transmitir la misión social de las ESAL, lo que hace necesario el desarrollo de nuevos modelos. Esto resalta la importancia de la transparencia en el sector sin fines de lucro, pero evidencia la dificultad de medirla empíricamente en la literatura anterior. Además, proponemos una revisión bibliográfica detallada que destaca el escaso número de artículos empíricos sobre medición de transparencia. Desde el punto de vista empírico, aplicamos el novedoso método Best-Worst para cubrir esta brecha. Nuestro objetivo último es que esta investigación sea útil para la comunidad científica, las autoridades públicas reguladoras o las autoridades supervisoras del dinero empleado por las ESAL.

**Palabras clave:** Transparencia, indicadores, organizaciones sin ánimo de lucro (ESAL), Método Best-Worst (BWM), toma de decisiones con criterios múltiples.

**Referencias:**

Rezaei, J. (2015). Best-worst multi-criteria decision-making method. *Omega (United Kingdom)*, 53, 49-57. <https://doi.org/10.1016/j.omega.2014.11.009>

# ***La estadística como herramienta de búsqueda de genes candidatos en los trastornos del espectro autista***

Leticia Díaz-Beltrán<sup>(1,2)</sup>, Francisco J. Esteban<sup>(2)</sup>

(1) UGC Oncología Médica, Hospital Universitario de Jaén, e-mail: ldiaz@fibaosalud.com

(2) Biología Experimental, Universidad de Jaén, e-mail: ldiazbel@ujaen.es / festeban@ujaen.es

## **RESUMEN**

Los trastornos del espectro autista (TEA) son un grupo de trastornos del desarrollo que afectan la comunicación y el comportamiento y cuyos síntomas generalmente aparecen durante los primeros dos años de vida. Las personas con TEA presentan dificultad para comunicarse e interactuar con otras personas, intereses limitados y comportamientos repetitivos o estereotipados. Se utiliza el término “espectro” para describir el autismo porque existe una amplia variación en el tipo y la gravedad de los síntomas que experimentan las personas con TEA, además de una gran comorbilidad con otros trastornos y enfermedades.

En la actualidad, aún no se sabe con certeza cuáles son las causas exactas del autismo, pero sí que está claro que resulta de la combinación de factores genéticos, neurológicos, inmunológicos y ambientales, y es por ello que se hace necesaria una aproximación multidisciplinar para estudiar el origen de estos trastornos.

Los recientes avances en genética, genómica, neurobiología del desarrollo y biología de sistemas han permitido entender el flujo de información desde los genes a las estructuras moleculares, su función bioquímica, conducta biológica y, finalmente, su influencia en los mecanismos biológicos responsables de los TEA. Las técnicas de microarrays y los análisis de secuenciación masiva generan gran cantidad de información biológica que necesita ser ordenada, procesada, analizada e interpretada con el fin de obtener resultados que puedan ser utilizados para dar respuesta a las demandas surgidas de la medicina molecular y la asistencia sanitaria personalizada. Es en este procesamiento y análisis de datos biológicos donde la estadística juega un papel fundamental, ya que nos permite establecer un sistema organizado de investigación, desde el diseño del experimento, sistema de muestreo, el control de calidad y el análisis hasta la presentación de la información.

Con este planteamiento, hemos optimizado un método estadístico basado en el análisis comparativo e integrado de diversos experimentos de expresión génica en autismo, tanto en sangre como en cerebro, para la búsqueda de genes candidatos que puedan estar involucrados en estos trastornos. Nuestro objetivo era determinar si existe una huella biológica común representativa de los TEA y probar si puede ser detectada tanto en sangre como en cerebro.

Para ello, recopilamos un set completo de 9 experimentos de expresión génica en autismo disponibles, de libre acceso y ya publicados, y obtuvimos 27 biosets o casos experimentales que proporcionaban listas de genes diferencialmente expresados en estos trastornos. El siguiente paso fue convertir estas listas de genes en matrices binarias de presencia/ausencia de genes con respecto a cada caso experimental (bioset).

Mediante un análisis cluster, utilizamos el Coeficiente de Jaccard para generar un dendrograma basado en genes para cada uno de los biosets, con el objetivo de establecer una relación de parentesco entre ellos. El estadístico de Jaccard mide el grado de similitud y diversidad entre muestras y se define como la cardinalidad de la intersección de ambos conjuntos dividida por la cardinalidad de su unión; es decir, es igual a 0 si no hay genes que intercepten e igual a 1 si todos los genes interceptan. El motivo por el que decidimos utilizar este coeficiente de distancia es porque está originalmente desarrollado para descubrir patrones en matrices binarias y no considera la ausencia compartida de una característica como evidencia de la existencia de una relación, algo fundamental en el contexto de la identificación de similitud entre experimentos de expresión génica. Una vez obtenido el dendrograma, validamos nuestro análisis cluster mediante un procedimiento de bootstrapping no paramétrico de muestreo sin reemplazo, con el objetivo de asegurar la estabilidad y validez de los subgrupos del dendrograma. Los subgrupos o clusters que presentaban un coeficiente de Jaccard  $> 0.6$  fueron considerados robustos y estables, mientras que los que tenían un valor  $\leq 0.5$  fueron descartados para el análisis.

Tal y como era de esperar, la mayoría de los 27 casos experimentales fueron agrupados primero por experimento y segundo por el tipo de tejido (sangre o cerebro). Sin embargo, un subgrupo se desviaba de lo esperado, agrupando biosets de tres experimentos diferentes e implicando ambos tipos de tejido, cerebro y sangre. Dicho cluster estaba conformado por un conjunto de 66 genes validados estadísticamente como significativos por nuestro análisis bootstrap.

Para evaluar la hipótesis de que estos 66 genes podían representar una huella biológica común para los TEA tanto en sangre como en cerebro, examinamos la red biológica que conformaban estos genes, su implicación en procesos biológicos y en otros trastornos neurológicos relacionados con autismo y su carga mutacional, revisando los más importantes análisis de secuenciación de exomas realizados hasta la fecha.

Como resultado, obtuvimos que 4 genes de nuestra lista de 66 candidatos, ya habían sido asociados a los TEA; además, 32 interactuaban directamente con genes que habían sido relacionados con autismo. Otros 19 genes de este grupo de 66 candidatos estaban directamente involucrados o presentaban vínculos con genes asociados a condiciones neurológicas relacionadas con los TEA, incluyendo esquizofrenia, epilepsia, discapacidad intelectual, déficit de atención y trastornos de conducta, Síndrome de Angelman, trastorno bipolar y retraso mental, entre otros. Triangulando con estudios de secuenciación de exomas, 9 de los 66 genes candidatos tenían variantes reportadas como de novo o riesgo elevado para autismo. En cuanto a su implicación en rutas biológicas, confirmamos que los 66 genes conformaban una red biológica muy interconectada y enriquecida en procesos biológicos relacionados con la transmisión sináptica, neurodegeneración, morfología anormal del cerebro, aprendizaje y memoria, apoyando nuestra hipótesis de que estos 66 genes candidatos podrían ejercer un papel relevante en la etiología del autismo.

En conclusión, el manejo de herramientas de análisis estadístico es hoy en día imprescindible para procesar, analizar y poder interpretar la ingente cantidad de datos biológicos obtenidos a través de las nuevas técnicas de secuenciación masiva, sobre todo en

desórdenes con una compleja arquitectura genética, como son los trastornos del espectro autista. Gracias a la aplicación de técnicas estadísticas en combinación con aproximaciones propias de la biología de sistemas, se puede progresar en el descubrimiento y validación de nuevas dianas terapéuticas que permitan esclarecer los mecanismos moleculares implicados en los TEA y desarrollar e implementar una medicina de precisión para estos trastornos.

**Palabras clave:** Análisis de agrupamiento, bootstrap, dendrograma, exomas, Jaccard, microarray.

## ABSTRACT

Autism spectrum disorders (ASD) are a group of developmental conditions that affect communication and behaviour, whose symptoms usually appear during the first two years of life. Autistic individuals have difficulty communicating and interacting with other people, limited interests, and repetitive or stereotyped behaviors. The term "spectrum" is used to describe autism because there is a wide variation in the type and severity of symptoms experienced by people with ASD, as well as a high degree of co-morbidity with other disorders and diseases.

To date, the exact causes of autism are not yet known, but it is clear that it results from a combination of genetic, neurological, immunological and environmental factors, that is why a multidisciplinary approach is needed to study the origin of these disorders.

Recent advances in genetics, genomics, developmental neurobiology and systems biology have made it possible to understand the flow of information from genes to molecular structures, their biochemical function, biological behaviour and, finally, their influence on the biological mechanisms responsible for ASD. Microarray techniques and whole genome sequencing analyses generate a large amount of biological information that needs to be ordered, processed, analysed and interpreted in order to obtain results that can be used to respond to the demands arising from molecular medicine and personalised health care. It is in this processing and analysis of biological data where statistics play a fundamental role, since it allows us to establish an organized research system, from the design of the experiment, sample collection, quality control and analysis, to the presentation of the information.

With this aim, we have optimized a statistical method based on the comparative and integrated analysis of various gene expression experiments in autism, both in blood and brain, for the search of candidate genes that may be involved in these disorders. Our goal was to determine whether there is a common biological signature representative of ASD and to test whether it can be detected in both blood and brain.

For that purpose, we compiled a complete set of 9 published and publically available ASD gene expression experiments and derived lists of differentially expressed genes from 27 case-control biosets. We then converted the transcriptional probe lists into a matrix of binary gene presence/absence with respect to each experimental case. We analyzed the

matrix using the Jaccard coefficient to construct a gene-based dendrogram for all 27 biosets with the aim of determining the relatedness among them. This statistic measures the similarity and diversity among sample sets and is defined by the size of the intersection divided by the size of the union of sample sets; that is, it is equal to 0 if no genes intersect and equal to 1 if all genes intersect. We elected to use this distance coefficient as it was originally developed for pattern discovery with binary matrices and does not treat the shared absence of a characteristic as evidence for relatedness, a valuable characteristic in the context of identifying similarity across gene expression experiments. Once the dendrogram is obtained, we validated our cluster analysis by means of a non-parametric bootstrapping procedure of sampling without replacement, in order to assess the stability and validity of the dendrogram's subgroups. Clusters with a Jaccard coefficient  $> 0.6$  were considered robust and stable, while those with a value of  $\leq 0.5$  were discarded for analysis.

As expected, most of the 27 experimental cases were grouped first by experiment and second by tissue type (blood or brain). However, one subgroup of the tree deviated from this expectation, including biosets from three different experiments and involving both tissue type, brain and blood. This cluster was rooted by a set of 66 genes statistically validated as significant by our bootstrap analysis.

To evaluate the hypothesis that this Root 66 could represent a common signature in blood and brain indicative of autism, we examined the genetic network formed by these genes, their implication in other autism-related neurological conditions and their mutational burden as reported by recently published exome sequencing studies.

As a result, we found that 4 genes from our list of 66 candidates had already been associated with ASD; in addition, 32 interacted directly with genes that had been linked to autism. Another 19 genes from this group of 66 candidates were directly involved or linked with genes associated with ASD-related neurological conditions, including schizophrenia, epilepsy, intellectual disability, attention deficit and conduct disorders, Angelman syndrome, bipolar disorder, and mental retardation, among others. Triangulating with exome sequencing studies, 9 of the 66 candidate genes had variants reported as *de novo* or high risk for autism. Regarding their involvement in biological pathways, we confirmed that the 66 genes formed a highly interconnected network, enriched in biological processes related to synaptic transmission, neurodegeneration, abnormal brain morphology, learning and memory, supporting our hypothesis that these 66 candidate genes may play a relevant role in the etiology of autism.

In conclusion, the use of statistical analysis tools becomes essential to process, analyze and interpret the huge amount of biological data obtained through next generation sequencing techniques, especially in disorders with a complex genetic architecture, such as ASD. Thanks to the application of these statistical techniques, in combination with systems biology approaches, progress can be made in the discovery and validation of new therapeutic targets that make it possible to clarify the molecular mechanisms involved in autism and to develop and implement precision medicine for this condition.

**Keywords:** Cluster analysis, bootstrap, dendrogram, exomes, Jaccard, microarray.

# *Asociación entre genotipo y respuesta farmacológica en la encefalopatía epiléptica STXBP1*

Paula Gregorio Pablos<sup>(1)</sup>, Francisco J. Esteban<sup>(2)</sup>

(1) *Dep. Biología Experimental, Universidad de Jaén, e-mail: pgp00016@red.ujaen.es*

(2) *Dep. Biología Experimental, Universidad de Jaén, e-mail: festeban@ujaen.es*

## RESUMEN

Las encefalopatías epilépticas infantiles tempranas (EIEE, Orphan#1934) agrupan una serie de síndromes y enfermedades raras que se caracterizan por tres criterios diagnósticos principales: crisis médicamente refractarias, encefalopatía difusa y retraso en el desarrollo psicomotor severo. En esta clasificación se incluyen el Síndrome de Ohtahara; el de Dravet (SD); el de Lennox-Gastaut; el de espasmos epilépticos infantiles ligado al cromosoma-X, también conocido como Síndrome de West (SW); y la denominada como EIEE4 (OMIM #6121649), causada por mutaciones en el gen que codifica para la proteína de unión a sintaxina (STXBP1), también conocida como Munc18-1. Desgraciadamente, todas estas patologías están en una fase inicial de estudio, su diagnóstico genético está pobremente desarrollado y carecen de un tratamiento farmacológico eficaz, bien definido y sin efectos secundarios graves.

Conforme se han ido caracterizando los síndromes de epilepsia familiar, se ha observado una muy amplia gama de fenotipos en los portadores, incluso se pueden observar diferencias fenotípicas clave entre individuos con la misma alteración genética. Así pues, se puede decir que existen co-factores que influyen en la expresión fenotípica que aún no se han determinado, sin descartar la existencia de genes epileptogénicos que juegan un marcado papel. El tipo y momento de las mutaciones durante el desarrollo, el tiempo y la localización de la expresión génica fisiológica, los factores epigenéticos, y modificadores de genes son algunos de los co-factores que pueden estar contribuyendo a la heterogeneidad fenotípica en las encefalopatías epilépticas. Esta variabilidad es un tema emergente de estudio para las epilepsias genéticas, ya que su investigación tiene un alto impacto al permitir enfoques de medicina personalizada para estas condiciones.

En definitiva, podemos decir que: (i) el diagnóstico del síndrome STXBP1 está pobremente desarrollado; (ii) los diferentes tipos de mutaciones encontradas (genotipo) aún no están bien correlacionadas con las manifestaciones clínicas de cada individuo (fenotipo); (iii) diferentes pacientes con una misma mutación pueden mostrar variabilidad en los síntomas o en el grado de afectación; (iv) individuos con una sintomatología similar pueden ser portadores de un tipo diferente de mutación; y (v) junto a la variabilidad genotípica y fenotípica de cada paciente hay que añadir la diferente respuesta que presentan al tratamiento. Como resultado, todo ello conlleva la falta de un correcto consejo genético y asistencia a los padres, su alarma y un retraso sustancial en la correcta aplicación de tratamientos adecuados, lo que repercute gravemente en el pronóstico de estos niños. Para solucionar esta situación proponemos realizar un análisis computacional desde una perspectiva de Biomedicina de Sistemas, es decir, mediante la integración y el análisis de datos masivos, clínicos y moleculares, a distintos niveles utilizando principalmente métodos computacionales. Para ello, accedemos a información biomédica, genómica y farmacológica

recogida en bases de datos públicas. Con los datos obtenidos desarrollamos predictores computacionales de asociación entre genotipo y respuesta farmacológica. Los análisis se llevan a cabo con el uso de WEKA ([cs.waikato.ac.nz/ml/weka](http://cs.waikato.ac.nz/ml/weka)) y scripts personalizados en Python ([python.org](http://python.org)). La idoneidad de clasificación y predicción se establece a través de la información proporcionada por matrices de confusión, área bajo curvas ROC y estadístico kappa, entre otros. De este modo, pretendemos contribuir al desarrollo del conocimiento del síndrome STXBP1 y permitir el desarrollo de una medicina personalizada en la encefalopatía epiléptica infantil.

**Palabras clave:** Análisis masivos de datos, Biomedicina de sistemas, respuesta farmacología, Munc18-1.

## ABSTRACT

Early-onset childhood epileptic encephalopathies (EIEE, Orphan #1934) is a group of syndromes and rare diseases that are characterized by three main diagnostic criteria: medically refractory seizures, diffuse encephalopathy and a severe delay in the psychomotor development. This classification includes Ohtahara Syndrome, Dravet (SD), Lennox-Gastaut, infantile epileptic spasms linked to the X-chromosome, also known as West Syndrome (SW), and the so-called EIEE4 (OMIM # 6121649), caused by mutations in the gene that codes for the syntaxin binding protein (STXBP1), also known as Munc18-1. Unfortunately, all these pathologies are in an initial phase of study, their diagnosis is poorly developed and they lack an effective, well-defined pharmacological treatment without serious side effects.

As family epilepsy syndromes have been characterized, a wide range of phenotypes have been observed in carriers. Moreover, key phenotypic differences between individuals with the same genetic alteration can be observed. Thus, it can be said that there are cofactors influencing phenotypes, not yet characterized, and without ruling out the existence of epileptogenic genes that play a key role.

The type and timing of mutations during development, time and location of physiological gene expression, epigenetic factors, and gene modifiers are some of the co-factors that may be contributing to phenotypic heterogeneity in epileptic encephalopathies. This variability is an emerging topic of study for genetic epilepsies, since their research has a high impact, mainly allowing personalized medicine approaches to these conditions. In short, we can say that: (i) the diagnosis of STXBP1 syndrome is poorly developed; (ii) the different types of mutations found (genotype) are not yet well correlated with the clinical manifestations of each individual (phenotype); (iii) different patients with the same mutation can show variability in symptoms or in the degree of affectation; (iv) individuals with a similar symptomatology can be carriers of a different type of mutation; and (v) together with the genotypic and phenotypic variability of each patient, the different response to the treatment must be added. As a result, all this entails the lack of proper genetic counseling and assistance to parents, their alarm and a substantial delay in the correct application of appropriate treatments, which has a serious impact on the prognosis of these children. To solve this situation, we propose to carry out a Systems Biomedicine approach, i.e., the integration and analysis of massive, clinical and molecular data, at different levels using

mainly computational methods. Biomedical, genomic and pharmacological is collected from public databases.

With the data obtained, we develop computational predictors of association between genotype and pharmacological responses. The analyses are carried out with the use of WEKA ([cs.waikato.ac.nz/ml/weka](http://cs.waikato.ac.nz/ml/weka)) and custom scripts in Python ([python.org](http://python.org)). The accuracy of classification and prediction is established through the information provided due to confusion matrices, area under ROC curves and kappa statistics, among others. In this way, we intend to contribute to the development of knowledge of STXBP1 syndrome and allow the development of a personalized medicine in childhood epileptic encephalopathies.

**Keywords:** Massive data analysis, Systems Biomedicine, pharmacological response, Munc18-1.

# *Tratamiento de la multicolinealidad en el modelo CAMELS a través de la regresión Cresta*

Ainara Rodríguez Sánchez<sup>(1)</sup>, Román Salmerón Gómez<sup>(2)</sup>, Catalina García García<sup>(3)</sup>

(1) Departamento de Economía, Universidad de Jaén, e-mail: [arsanche@ujaen.es](mailto:arsanche@ujaen.es)

(2) y (3) Departamento de Métodos Cuantitativos para la Economía y la Empresa, Universidad de Granada, e-mail: [romansg@ugr.es](mailto:romansg@ugr.es) y [cbgarcia@ugr.es](mailto:cbgarcia@ugr.es)

## RESUMEN

El modelo CAMELS es utilizado para evaluar el riesgo de las distintas entidades de crédito, tema muy importante en la actualidad. A menudo las variables de esta metodología son utilizadas como variables independientes de un modelo de regresión lineal, las cuáles pueden estar relacionadas entre sí dando lugar a un problema denominado multicolinealidad. Si se obvia este problema y se estima el modelo por Mínimos Cuadrados Ordinarios pueden aparecer estimaciones inestables entre otras consecuencias. En este trabajo se pretende tratar dicho problema a través de un método alternativo de estimación muy conocido denominado regresión Cresta, utilizando para ello datos reales de grandes entidades financieras a nivel internacional.

**Palabras clave:** CAMELS, riesgo, regresión lineal, multicolinealidad, regresión Cresta.

## ABSTRACT

The CAMELS model is used for evaluating the risk of the different banks, topic very important in the nowadays. The variables of this model commonly used as independents variables in lineal regression model leading to multicollinearity since these variables may be related between them. If this problem is disregarded and the model is estimated by Ordinary Least Squares is obtained instable estimations and other consequences. The intention in this paper is to treatment this problem with a alternative estimation method very well-known called Ridge regression, using real data from large international financial institutions.

**Keywords:** CAMELS, risk, lineal regression, multicollinearity, Ridge regression.

***Valor de la transparencia en entidades sin ánimo de lucro:  
propuesta de mejora de una batería de indicadores***  
Cristina Ortega-Rodríguez<sup>(1)</sup>, Antonio Luis Moreno Albarracín<sup>(2)</sup>, Ana Licerán-  
Gutierrez<sup>(3)</sup>, Álvaro Labella Romero<sup>(4)</sup>

(5) Departamento de Economía Financiera y Contabilidad: [corodrig@ujaen.es](mailto:corodrig@ujaen.es)

(6) Departamento de Economía Financiera y Contabilidad: [almoreno@ujaen.es](mailto:almoreno@ujaen.es)

(7) Departamento de Economía Financiera y Contabilidad; [aliceran@ujaen.es](mailto:aliceran@ujaen.es)

(8) Departamento de Informática; [alabella@ujaen.es](mailto:alabella@ujaen.es)

## INTRODUCCIÓN

Las entidades sin ánimo de lucro (ESAL, en adelante) son cruciales debido a la actividad social que desarrollan en beneficio de las personas más necesitadas de la sociedad. Para desarrollar su actividad reciben recursos de entidades públicas y privadas, por lo que es necesario una evaluación continua de que la gestión de los fondos recibidos se está empleando adecuadamente. En estos términos, la transparencia de la información revelada y las prácticas de buen gobierno presentan un interés creciente.

## OBJETIVOS

De la información obtenida de las aportaciones estudiadas hasta la fecha, se deduce que, la implantación de indicadores en el ámbito de las ESAL es un campo de reciente aparición y apenas desarrollado. El **objetivo general** de esta tesis doctoral es establecer un sistema de medición de la credibilidad de la información divulgada por las ESAL que tenga en cuenta los diferentes aspectos señalados por la literatura como representativos de transparencia y de buen gobierno. Haciendo uso de indicadores ya definidos, se pretende ofrecer una adecuada ponderación de los mismos validado empíricamente mediante la aplicación de técnicas estadísticas rigurosas que, además, estén basadas en la opinión de expertos en la divulgación de información. Con todo ello, se propondrá un sistema capaz de valorar apropiadamente desde el punto de vista cuantitativo el grado de transparencia y buen gobierno de la información divulgada por las ESAL.

## METODOLOGÍA

Se proponen varios **métodos de investigación**. En concreto, serían los siguientes:

- Realizar una revisión sistemática de la literatura sobre la transparencia y buen gobierno de las ESAL. De la revisión se pretende definir claramente cuál es el gap de investigación a cubrir con nuestro presente trabajo.
- Aplicar el proceso de selección multicriterio del Best-Worst Method (BWM, en adelante) basado en la opinión de expertos, que se obtendrá mediante la realización de una serie de cuestionarios sobre qué aspectos consideran dichos expertos como más relevantes para representar el grado de transparencia y de buen gobierno de la información divulgada por las ESAL. Con el software específico, se espera poder definir una ponderación adecuada de los múltiples indicadores propuestos.
- Se utilizará la técnica del panel de expertos. A través del panel de discusión con varios expertos en el ámbito de las ESAL expresarán sus opiniones sobre qué entender por

transparencia y buen gobierno, qué aspectos consideran más relevantes y menos, etc. Así, se pretende enriquecer el análisis cuantitativo con un conocimiento más profundo del objeto de estudio del presente trabajo de investigación.

- Fase final: Desarrollo de un sistema inteligente que ayude a la decisión para clasificar la transparencia en las Entidades Sin Ánimo de Lucro.

## RESULTADOS

Una **primera fase de revisión de literatura** permite evidenciar una serie de problemas respecto a los indicadores referidos a la credibilidad de las ESAL:

- Si bien transparencia y buen gobierno están ampliamente tratados desde el punto de vista teórico en literatura, no hay muchos estudios que los analicen empíricamente;
- Resulta necesario cuantificar el grado de transparencia y buen gobierno de las ESAL, pero el uso de indicadores, presenta diversos problemas: I) Existen diversas baterías de indicadores de transparencia y de buen gobierno, pero cada batería considera múltiples aspectos, sin existir un criterio unánime sobre cuáles los representan mejor; II) Dichas baterías no especifican el peso que cada indicador; III) La única batería que proporciona pesos para los indicadores, es la de la Coordinadora de ONGD de España (CONGDE, en adelante) (2019) que, ofrece una ponderación subjetiva.

En una **segunda fase empírica**, nuestros resultados evidencian diferencias interesantes en la importancia de los indicadores propuestos por expertos, lo que sugiere una necesidad de revisión de la importancia propuesta por CONGDE (2019) para los indicadores.

## CONCLUSIÓN

Surge la necesidad de implantar un conjunto de indicadores que nos permitan complementar la información suministrada por los estados contables tradicionales, estando dichos indicadores adecuadamente ponderados en consonancia con la opinión de expertos en el ámbito de las ESAL. El propósito de esta investigación responde a la necesidad de una adecuada rendición de cuentas de las ESAL, marcada por un adecuado grado de transparencia de la información divulgada por dichas entidades, de cara a poder justificar el empleo de los fondos recibidos tanto por entes públicos como por particulares para cumplir la misión social que les es encomendada.

**Palabras Clave:** Transparencia, buen gobierno, indicadores, Entidades sin ánimo de lucro (ESAL), método Best-Worst.

# *Simulación de un modelo de regresión de Poisson para datos de panel*

Alba Patricia Espinoza Rojas<sup>(1)</sup>, José Rodríguez Avi<sup>(1)</sup>

(1) Dpto. de Estadística e Investigación Operativa. Universidad de Jaén, España

## RESUMEN

El uso de modelos de datos de panel está ampliamente extendido en el contexto de ciencias económicas, entre otros. Este tipo de modelos surgen al analizar una variable en función de otras variables exógenas provenientes de datos tomados tanto longitudinalmente como transversalmente. En este tipo de modelos las variables exógenas pueden ser cuantitativas o cualitativas, mientras que lo más frecuente para la variable endógena o explicada es que se considere cuantitativa y continua. De esta manera se supone que para cada combinación de valores de las variables exógenas, los residuos siguen una distribución normal.

Sin embargo, en múltiples ocasiones, la variable objeto de estudio es una variable discreta, más concretamente es una variable de conteo, por lo que toma valores dentro de los números naturales. En estos casos la suposición de normalidad en los residuos no es válida, sino que se ha de asumir que la distribución residual es una distribución de conteo, como por ejemplo, una distribución de Poisson en la que la media (el parámetro) se obtiene en función de las variables exógenas. En este trabajo presentamos un estudio de simulación para este tipo de modelos en el que, bajo condiciones controladas de las variables exógenas se simulan valores de la variable de respuesta a través de números aleatorios procedentes de la correspondiente distribución de Poisson, lo que nos permite analizar el procedimiento de obtención de los coeficientes del modelo y se estudian sus propiedades estadísticas, para verificar las estimaciones obtenidas.

**Palabras clave:** datos de panel, distribuciones de conteo, distribución de Poisson

# *Lectura de tablas estadísticas por futuras maestras de Educación Infantil en Chile*

Danilo Díaz-Levicoy<sup>(1)</sup>, Karen Ruiz<sup>(2)</sup>, Carmen Gloria Aguayo<sup>(3)</sup>, Valeria Bizet<sup>(4)</sup>, Roxana Alemán<sup>(5)</sup>

(1) Departamento de Matemática, Física y Estadística, Universidad Católica del Maule, dddiaz01@hotmail.com

(2) Estudiante de doctorado, Universidad de Granada, karenruizreyes@gmail.com

(3) Departamento de Didáctica de la Matemática, Universidad de Granada, cgaguayo@ugr.es

(4) Estudiante de doctorado, Universidad de Granada, valeriabizet@gmail.com

(5) CIEPLANE, Universidad Autónoma "Juan Misael Saracho", raleman@correo.ugr.es

## RESUMEN

En este trabajo mostramos los resultados de un estudio sobre los niveles de lectura de tablas estadísticas por futuras maestras de Educación Infantil de una universidad chilena, mediante la aplicación de un instrumento previamente validado por expertos. Para obtener la información se realizó un análisis de contenido a las respuestas entregadas por la muestra de futuras maestras. Los resultados evidencian que se encuentran, mayoritariamente, en los niveles leer los datos y leer dentro los datos, los que están asociados a la lectura literal y al desarrollo de procesos matemáticos sencillos de la información mostrada en la tabla.

**Palabras clave:** tablas estadísticas, niveles de lectura, futuras maestras, Educación Infantil.

## ABSTRACT

In this work we show the results of a study on the levels of reading statistical tables by future teachers of Early Childhood Education of a Chilean university, through the application of an instrument previously validated by experts. To obtain the information, a content analysis was performed on the responses delivered by the sample of future teachers. The results show that, for the most part, the levels read the data and read the data, which are associated with the literal reading and the development of simple mathematical processes of the information shown in the table.

**Keywords:** statistical tables, reading levels, prospective teachers, Early Childhood Education.

## *Estudio de ciertas falacias cometidas en los medios de comunicación*

Francisco Martínez Ortiz<sup>(1)</sup>, Felipe Ruz<sup>(2)</sup>, Danilo Díaz-Levicoy<sup>(3)</sup>, Karen Ruiz<sup>(4)</sup>, Elena Molina-Portillo<sup>(5)</sup>

(1) Departamento de Didáctica de las Ciencias, Universidad de Jaén, fmortiz@uja.es

(2) Estudiante de doctorado, Universidad de Granada, feliperuzangel@gmail.com

(3) Departamento de Matemática, Física y Estadística, Universidad Católica del Maule, dddiaz01@hotmail.com

(4) Estudiante de doctorado, Universidad de Granada, karenruizreyes@gmail.com

(5) Departamento de Didáctica de la Matemática, Universidad de Granada, elemo@ugr.es

### RESUMEN

Los medios de comunicación, a menudo, presentan noticias que provocan que los consumidores de datos no realicen una correcta interpretación de la información que se está transmitiendo, a pesar de que la persona que la recibe no es consciente de ello. Por tanto, es esencial que los receptores de la información tengan una postura crítica adecuada para no cometer ciertas falacias a la hora de realizar una conclusión adecuada sobre estas noticias.

La Real Academia Española define falacia como “el engaño, fraude o mentira con que se intenta dañar a alguien o algo”. Algunos autores indican que una falacia es “una mentira disfrazada de verdad”, mientras que otros argumentan que un juicio de razonamiento es denominado falacia cuando una gran parte de la gente que la comete, después de una explicación razonable, está dispuesta a aceptar que no es un error trivial, por lo que se repetiría en razonamientos similares. Además, el error es conceptual, no meramente verbal o técnico y las personas que lo han cometido deberían haber sabido la respuesta o un procedimiento para encontrarla.

Una de las falacias que más se produce en los receptores de cierta información estadística es la *falacia de la base extensa*, que consiste en impresionar con la gravedad de la situación al hablar de un fenómeno raro que afecte a una base amplia de la población, utilizando sólo parte de la información o sin tener en cuenta alguna información relacionada con los datos. Se sigue la estrategia de hablar de números absolutos y no de porcentajes (o viceversa), es decir, se intenta impresionar al lector utilizando un número absoluto en un caso en que si se utilizase un porcentaje la noticia no tendría tanto impacto. Un ejemplo de ello se muestra en la Figura 1.



Figura 1: Noticia de elPeriódico.com relacionada con la falacia de la base extensa

Otra de las falacias más comunes es la *falacia de las tasas base*. Aparece cuando el receptor ignora la información general en un evento, guiándose únicamente por la información específica. Al igual que en otras falacias como la *falacia de la base extensa*, se omite una parte de la información que es importante para llegar a comprender la noticia. Un ejemplo relacionado con la *falacia de las tasas base* es el que puede observarse en la Figura 2.



Figura 2: Noticia de Antena 3 relacionada con la falacia de las tasas base

Por último, la *falacia de las comparaciones en valor absoluto* trata de impresionar al lector, en este caso, utilizando valores absolutos en vez de tasas al comparar dos poblaciones mostrando sólo parte de la información. Esta falacia es bastante común en los medios de comunicación y en la vida cotidiana de los ciudadanos. Por ello, es importante que el receptor sea capaz de abstraerse del razonamiento propio de cualquier ser humano, es decir, aquel que hace que se asuma como cierta la información que se está recibiendo. Sin embargo, este proceso es bastante complicado y contra intuitivo. En la Figura 3 puede observarse un ejemplo de ella



Figura 3: Noticia de Diario de Sevilla relacionada con la falacia de las comparaciones en valor absoluto

**Palabras clave:** falacia, información estadística, medios de comunicación

## *Comprensión de la selección de muestras*

Karen Ruiz<sup>(1)</sup>, Felipe Ruz<sup>(2)</sup>, Francisco Martínez Ortiz<sup>(3)</sup>, Carmen Gloria Aguayo<sup>(4)</sup>, José Miguel Contreras<sup>(5)</sup>

(1) Estudiante de doctorado, Universidad de Granada, karenruizreyes@gmail.com

(2) Estudiante de doctorado, Universidad de Granada, feliperuzangel@gmail.com

(3) Departamento de Didáctica de las Ciencias, Universidad de Jaén, fmortiz@uja.es

(4) Departamento de Didáctica de la Matemática, Universidad de Granada, cgaguayo@ugr.es

(5) Departamento de Didáctica de la Matemática, Universidad de Granada, jmcontreras@ugr.es

### RESUMEN

Este trabajo tiene por objetivo indagar qué comprenden los estudiantes chilenos de secundaria sobre métodos de selección de muestras y si son capaces de reconocer el sesgo en diferentes muestras obtenidas. Se presenta la síntesis del análisis de las respuestas de 1241 estudiantes chilenos de secundaria, a un Ítem de respuesta abierta.

Se incluyen tres métodos de muestreo básicos en este Ítem: muestreo aleatorio, restringido y autoseleccionado. El método de *muestreo aleatorio* dio a cada miembro de la población la misma oportunidad de ser seleccionados. Los métodos de *muestreo restringidos* preguntaron a grupos particulares de personas que podían ser más capaces de seleccionar una respuesta determinada y, en consecuencia, sesgar los resultados en una dirección particular. Los métodos de *muestreo autoseleccionados* permitieron que los participantes se seleccionaran ellos mismos y fueran problemáticos porque no había suficiente evidencia para evaluar si la muestra era representativa (Jacobs, 1997).

Los estudiantes se inclinan principalmente a preferir métodos de muestreo estratificado por sobre a otros, ya que buscan que todos los sujetos tengan la misma posibilidad de participar, es decir, consideran que este método es más equitativo y justo.

Cuando los estudiantes se enfrentan a diferentes métodos de muestreo empleando encuestas, les dificulta identificar las ventajas del muestreo aleatorio, pues consideran que en el muestreo aleatorio no si es un método apropiado porque podría escoger a las personas equivocadas, es decir, no consideran que la importancia de la aleatoriedad en el método de selección de una muestra. Además, son incapaces de detectar los sesgos en los muestreos autoseleccionados, es decir, consideran apropiado que sólo se consideren las respuestas de quienes han contestado un cuestionario rápidamente.

**Palabras clave:** comprensión, educación estadística, muestra, selección de muestras.

## ANEXO

**Ítem.** Los estudiantes de un colegio realizaron una encuesta para determinar el porcentaje de niños que reciclan en sus casas:

- Andrea preguntó a 80 estudiantes que son miembros del club de medioambiente.
  - Elena envió un cuestionario a todos los niños del colegio y tomó los primeros 80 que contestaron.
  - Pedro quiso el mismo número de niños y niñas. Así que preguntó a 5 niños y 5 niñas de cada curso para conseguir los 80 estudiantes que constituyen la muestra.
  - Luis tenía los nombres de los 800 estudiantes en la escuela; puso cada nombre en un papel en un sombrero y sacó 80.
  - Sara no conocía demasiados niños por lo que decidió encuestar a 80 niñas. Pero quería asegurarse de tener niñas de varias edades, por lo que tomó 10 niñas de cada nivel.
- a. Completa la última columna donde preguntamos si el método de cada niño es apropiado.

Niño	A quien pregunta	¿Te parece un método apropiado? ¿Por qué?
Andrea	80 estudiantes del club de medioambiente.	
Elena	Envío un cuestionario a todos los niños del colegio y tomó los 80 que contestaron primero.	
Pedro	5 niños y 5 niñas de cada curso	
Luis	Puso el nombre de cada niño de la escuela (800) en un papel en un sombrero y sacó 80	
Sara	10 niñas de cada nivel.	

- b. De acuerdo a los datos entregados en el enunciado, ¿Cuál crees que es la mejor manera de escoger a los niños para estimar el porcentaje de ellos que recicla en casa?

# *Algunas implicaciones cívicas de la estadística: El caso del proceso electoral en España*

Felipe Ruz<sup>(1)</sup>, Elena Molina-Portillo<sup>(2)</sup>, Valeria Bizet<sup>(3)</sup>, Carmen Gloria Aguayo<sup>(4)</sup>, José M. Contreras<sup>(5)</sup>

*Departamento de didáctica de la matemática, Universidad de Granada*

(1) felipe.ruz.angel@gmail.com; (2) elemo@ugr.es; (3) valeriabizet@gmail.com; (4) cgaguayo@ugr.es; (5) jmcontreras@ugr.es;

## RESUMEN

Actualmente, garantizar la participación activa de los ciudadanos en debates o discusiones sobre aspectos de relevancia social, tales como el crecimiento poblacional, la desigualdad económica o la pobreza, requieren la comprensión, interpretación y evaluación crítica de estadísticas sobre tendencias pasadas, situaciones vigentes y posibles cambios futuros. Más aun, cuando la información sobre estos temas está a disposición de la ciudadanía a través de diversos medios de comunicación tradicionales y/o digitales. Este hecho requiere no solo dominar técnicas estadísticas, sino que implica generar hábitos mentales, motivados por el compromiso de interactuar con datos reales y comunicar evidencia numérica.

Por tanto, en este trabajo exponemos algunas implicaciones para la ciudadanía al enfrentarse a fenómenos de interés social como las elecciones parlamentarias, desde la perspectiva de las denominadas *estadísticas cívicas*. Dicha noción se conceptualiza como una sub-disciplina de la alfabetización o cultura estadística, para la cual se ha desarrollado un modelo teórico cuyo eje central es la comprensión de la sociedad y sus políticas, para el cual cobran sentido las estadísticas cívicas. Su estructura conceptual se organiza en una primera dimensión sobre aspectos afectivos (compromiso y acción) que se complementa con una segunda sobre cuestiones cognitivas (conocimiento) y una tercera sobre las competencias de comunicación involucradas (procesos habilitadores). En este contexto, dada la inestabilidad política que atravesó España hace un tiempo, nos propusimos identificar situaciones donde sea necesario aplicar destrezas estadísticas en el marco de las elecciones parlamentarias, sistematizadas según el modelo teórico de estadísticas cívicas descrito.

Entre los resultados, sobre la dimensión afectiva destacamos que usualmente las encuestas electorales presentan sesgos en función de quienes las financian y las realizan. Por tanto, es necesario cuestionar las afirmaciones que de ellas se deriven adoptando una postura crítica y reflexiva. En cuanto a la dimensión cognitiva, al abordar el problema de la estimación de voto previo a la celebración de las elecciones, destacamos la importancia de la elección de un modelo de predicción. Este hecho se evidencia al contrastar los resultados de las elecciones con las predicciones realizadas por diferentes entidades, como el Centro de Investigaciones Sociológicas (CIS) que emplea el Modelo V89 y no acierta en sus predicciones. Mientras que quienes utilizan el modelo desarrollado por Kiko Llaneras, basado en encuestas, precisión histórica y simulaciones de elecciones, estiman resultados muy cercanos a la realidad. En la dimensión sobre las competencias de comunicación, destacamos que las agencias nacionales e internacionales de estadística ofrecen sus datos

públicamente, pero el acceso a ellos involucra el uso de aplicaciones o recursos digitales que requieren habilidades específicas sobre TICs.

De esta forma, el análisis realizado destaca la urgencia de incorporar conocimientos estadísticos especializados durante la enseñanza obligatoria y, por ende, en la formación y desarrollo profesional de los profesores que tienen la responsabilidad de enseñarlos.

**Palabras clave:** Estadísticas cívicas, Elecciones políticas, Ciudadanía, España

## ABSTRACT

Currently, guaranteeing an active participation of citizens in debates and discussions on aspects of social relevance, such population growth, economic inequality or poverty, require the understanding, interpretation, and critical evaluation of statistics on past trends and possible future changes. Moreover, when information on these issues is available to citizens through media. This fact requires not only mastering statistical techniques, but also implies generating mental habits, motivated by the commitment to interact with real data, and communicate numerical evidence.

Therefore, in this work we expose some implications for citizenship when facing phenomena of social interest such as parliamentary elections, from the perspective of the so-called *civic statistics*. This notion is conceptualized as a sub-discipline of statistical literacy, for which a theoretical model has been developed, whose central topic is the understanding of society and its policies, where civic statistics make sense. The framework is organized in a first dimension on affective aspects (commitment and action) that is complemented by a second on cognitive issues (knowledge), and a third on the communication skills involved (enabling processes). In this context, given the political instability that Spain went through some time ago, we set out to identify situations where it is necessary to apply statistical skills within the parliamentary elections, systematized according to the civic statistics theoretical model described above.

Among the results, on the affective dimension, we highlight that usually the electoral surveys have biases depending on who finances and carries them out. Therefore, it is necessary to question the statements that are derived from them by adopting a critical and reflective position. About the cognitive dimension, when addressing the problem of estimating the pre-election vote, we highlight the importance of choosing a prediction model. This fact is evidenced by comparing the elections' results with the predictions made by different entities, such the CIS that uses the V89 Model and is not correct in its predictions. While those who use the model developed by Kiko Llaneras, based on surveys, historical accuracy and election simulations, estimate results very close to reality. In the communication skills dimension, we highlight that national and international statistical agencies offer their data publicly, but access to them involves the use of digital applications or resources that require specific ICT' skills.

The analysis carried out, highlight the importance of incorporating specialized statistical knowledge during compulsory education and, therefore, in the training and professional development of teachers who have the responsibility to teach statistics in school.

**Keywords:** Civic Statistics, Politician election, Citizenship, Spain.

# *Propuesta de un programa de intervención sobre problemas adaptativos asociados a la dependencia emocional*

Violeta Ruiz Sánchez

(1) Universidad de Jaén, e-mail: vrs00007@red.ujaen.es

## RESUMEN

Este estudio va dirigido a personas que carecen de control sobre su conducta, que quieren alejarse de alguien que le produce un daño físico y emocional y no precisa de “las estrategias adecuadas”. El trabajo aborda “La dependencia emocional” desde la perspectiva de diferentes autores, un marco contextual que nos sitúa en una época social en la que surge esta preocupación. En los siguientes capítulos seremos partícipes de una aproximación de un Programa de Intervención para proponer mejoras en las conductas de las personas y por consiguiente en su calidad de vida. Este programa va destinado a personas universitarias, lugar dónde se detectan más números de casos en dependencia en relación a su cónyuge al igual que su densidad poblacional nos proporciona datos más exhaustivos y representativos.

La intervención va acompañada de una breve mención sobre “la autoestima” ya que se considera un rasgo importante de medición. Pues, las personas con baja autoestima desarrollan mayor probabilidad de tener “dependencia emocional”. Para finalizar, indicamos las variables o características que desarrolla un “dependiente emocional” y abordamos una conclusión general con posibles soluciones (actividades).

**Palabras clave:** Intervención, dependencia emocional, autoestima baja, soluciones y universitarios.

## ABSTRACT

This study is aimed at people who lack control over their behavior, who want to get away from someone who causes physical and emotional damage and does not need "the right strategies". Proposal of Intervention in Emotional Dependency: We have decided to carry out this project because we believe it will be helpful for emotionally dependent people. Dependency on a partner, single people, relatives, objects, machines or drugs. Any state which gives rise to a yearning to see the desired stimulus again in a repetitive, disproportionate way, thus calming the longing for the presence of the said stimulus. The project centers on "emotional dependence" from the perspective of different authors, a contextual framework that situates us in the social era in which this concern arises, and the difference from the present. Next, we will examine an approach to a programme which can be applied to people at university. We will focus on a specific area, since the range is too wide for us to cover it entirely. This Intervention helps those who suffer from these problems, improving their quality of life and raising the awareness of minds which believe themselves to be problem-free. The intervention comes with a brief mention of "self-esteem", since it is considered a characteristic which is worth measuring, as people with low self-esteem have a greater probability of developing emotional dependence. Finally, we will allude to the variables or characteristics affecting those suffering from emotional dependency and we will draw a general conclusion with possible solutions.

**Keywords:** Intervention, emotional dependence, self-esteem, variables, programmer and solutions.

# *La Probabilidad en la enseñanza*

Pedro Antonio Cano Chica<sup>(1)</sup>

(1) Departamento de Estadística e Investigación Operativa, Universidad de Jaén, e-mail:  
[pacano@ujaen.es](mailto:pacano@ujaen.es)

## RESUMEN

El objetivo del presente trabajo es mostrar las competencias adquiridas durante la realización del Máster Universitario en Profesorado de Educación Secundaria Obligatoria y Bachillerato, Formación Profesional y Enseñanza de Idiomas. Este trabajo se fundamenta en el tema de Probabilidad y se dirige a los alumnos de 3º de Educación Secundaria Obligatoria. Para su elaboración, se ha llevado a cabo una fundamentación curricular, donde se ha realizado un estudio comparativo entre dos libros de texto y la normativa vigente. En segundo lugar, se ha efectuado una fundamentación epistemológica, donde se ha elaborado el tema número 63 del temario oficial de las oposiciones de la especialidad de Matemáticas. En tercer lugar, se ha llevado a cabo una fundamentación didáctica, donde se han analizado algunos artículos científicos en relación a las posibles dificultades y errores que el alumnado suele encontrar al enfrentarse a este contenido. Finalmente, se ha desarrollado una proyección didáctica del tema seleccionado.

**Palabras clave:** Matemáticas, Probabilidad, Estadística, unidad didáctica, aprendizaje, enseñanza, Educación Secundaria Obligatoria.

# *Una modelización para los centros de educación*

Valentina Cueva-López<sup>(1)</sup>, María José Olmo-Jiménez<sup>(2)</sup>, Julia Rodríguez Reinoso<sup>(3)</sup>

(1) Dpto. de Estadística e Investigación Operativa, Universidad de Jaén, España, [vcueva@ujaen.es](mailto:vcueva@ujaen.es)

(2) Dpto. de Estadística e Investigación Operativa, Universidad de Jaén, España, [miolmo@ujaen.es](mailto:miolmo@ujaen.es)

(3) Universidad Complutense de Madrid, España, [juliarodriguezreinoso@gmail.com](mailto:juliarodriguezreinoso@gmail.com)

## RESUMEN

La modelización estadística de variables que tengan baja frecuencia de 0, pero sí, grandes frecuencias en otros valores, se lleva a cabo a través de distribuciones que no son las más usuales, como la distribución de Poisson, sino por otras más recientes, como la distribución triparamétrica compleja de Pearson (CTP) o alguna de sus variantes. Por este motivo, vamos a proceder a la modelización de los centros educativos por municipios en distintas Comunidades Autónomas de España. Y veremos qué distribuciones, si la propuesta por este trabajo CTP o las propuestas por otros autores, son las más adecuadas, a través de los criterios de bondad de ajuste.

**Palabras clave:** Sobredispersión, Infradispersión Pearson, CTP, CBP, Educación

## ABSTRACT

Statistical modeling of variables that have low frequency of 0 but large frequencies in other values, is carried out through distributions that are not the most usual, such as the Poisson distribution, although it is possible to realize by more recent ones, such as Pearson's complex triparametric distribution (CTP) or some of its variants. For this reason, we will proceed to the modeling of educational centers by municipalities in different Autonomous Communities of Spain. And we will see which distributions, if the one proposed by this CTP work or those proposed by other authors, are the most appropriate, through the criteria of goodness of fit.

**Keywords:** Overdispersion, Pearson, CBP, Asturias