

LIBRO DE RESÚMENES

IV JORNADAS DE ESTADÍSTICA COMO HERRAMIENTA CIENTÍFICA

DEL 24 AL 26
MARZO 2021



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



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa y Escuela de Doctorado.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

ÍNDICE

SALUDO DEL PRESIDENTE	5
COMITÉ ORGANIZADOR	7
COMITÉ CIENTÍFICO	9
PROGRAMA	11
CIENCIAS DE LA SALUD	15
Bar-Lev, Bobovitch, and Boukai model to study sexual behavior	17
Beatriz Cobo ⁽¹⁾ , María del Mar Rueda ⁽²⁾ , Francisca López-Torrecillas ⁽³⁾	17
Modelos de elección discreta y consumo de tabaco	19
Imanol L. Nieto González	19
Propuesta de un modelo de salud mental considerando dimensiones psicológicas en tiempos de pandemia.	20
Armando González Sánchez ⁽¹⁾ , Greibin Villegas Barahona ⁽²⁾ M.Sc. Eva Carazo Vargas, ⁽³⁾ Máster Raúl Ortega ⁽⁴⁾ , Harold Arias LeClaire ⁽⁵⁾ .	20
Un algoritmo para estimar los parámetros de un biplot logístico aplicado a la caracterización de distintas variedades de <i>Colletotrichum graminicola</i>	21
Laura Vicente-Gonzalez ⁽¹⁾ , Michael R. Thon ⁽²⁾ , Serenella A. Sukno ⁽³⁾ & José L.Vicente-Villardón ⁽⁴⁾	21
CIENCIAS ECONÓMICAS	25
<i>Fuzzy Cognitive Maps</i> como herramienta aplicada en la toma de decisiones bajo alta incertidumbre	27
Julio Vena Oya ⁽¹⁾ , José Alberto Castañeda García ⁽²⁾ , Miguel Ángel Rodríguez Molina ⁽³⁾	27
Manuela Cortés ⁽¹⁾	29
Regresión logística aplicada a las finanzas del fútbol español	31
Alberto Calahorro López (1), Julio Vena Oya (2), Melinda Ratkai, PhD (3)	31
Análisis Estadístico de Datos Textuales sobre Estrategias de Comunicación Empresarial a Nivel Sectorial hacia el Compromiso con la Agenda 2030	34
Víctor Amor-Esteban ⁽¹⁾ , Isabel-María García-Sánchez ⁽²⁾ , David Galindo-Álvarez ⁽³⁾	34



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO
HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Una aproximación estadística de los grandes desequilibrios en el desarrollo socioeconómico del territorio portugués	36
Miguel Marques de Sousa	36
INGENIERÍA	39
Distribuciones Tipo Fase Lineales para la Modelización Probabilística de Datos Funcionales en Electrónica	41
Christian J. Acal González ⁽¹⁾ , Juan Eloy Ruiz Castro ⁽¹⁾ , Ana M. Aguilera del Pino ⁽¹⁾	41
Modelado estadístico de datos de error posicional. Un ejemplo práctico	43
Julia Rodríguez-Reinoso ⁽¹⁾ , José María Rodríguez-Reinoso ⁽²⁾	43
CIENCIAS SOCIALES	45
Análisis Multivariante de las políticas de cierre y contención efectivas frente al COVID-19: países de Europa	47
Joel Antonio Martínez-Regalado ⁽¹⁾ , Cinthia Leonora Murillo-Avalos ⁽²⁾ , Purificación Vicente-Galindo ⁽³⁾ , J.L. Vicente-Villardón ⁽⁴⁾	47
Cinthia L. Murillo-Avalos ⁽¹⁾ , Mitzi Cubilla-Montilla ⁽²⁾ , José Luis Vicente-Villardón ⁽³⁾ , Purificación Vicente-Galindo ⁽⁴⁾	49
Técnicas para reducir el sesgo de selección en una encuesta online sobre adicciones	51
Ramón Ferri-García ⁽¹⁾ , María del Mar Rueda ⁽²⁾ , Francisca López-Torrecillas ⁽³⁾	51
Análisis multivariante de la labor de la psicología en los Centros de Reconocimiento de Conductores	53
Armando González Sánchez ⁽¹⁾	53
Caracterización multivariante del uso de las TIC en estudiantes universitarios de Ciencias Sociales	54
María-Concepción Vega-Hernández ⁽¹⁾ , Carmen Patino-Alonso ⁽²⁾ , María-Purificación Galindo-Villardón ⁽³⁾	54
Caracterización de la desigualdad de género. Una perspectiva multivariante usando HJ-Biplot	57
Nerea González-García ⁽¹⁾ , Ana B. Nieto-Librero ⁽²⁾ , Purificación Galindo-Villardón ⁽³⁾	58
Estudio del Índice Desarrollo Humano a través del Clustering Disjoint HJ-Biplot	60
Ana B. Nieto-Librero ⁽¹⁾ , Nerea González-García ⁽²⁾ , Purificación Galindo-Villardón ⁽³⁾	60
Aplicación de técnicas estadísticas multivariantes y de reducción de dimensionalidad en variables de carga externa y contextuales en un equipo de fútbol semiprofesional.	62
Enrique Benítez Andrés ⁽¹⁾ , Mercedes Sanchez Barba ⁽²⁾ , Mario Sánchez García ⁽³⁾ , Javier Sánchez Sánchez ⁽⁴⁾	62
Using a Bayesian Probabilistic Model and Machine Learning Tools for Smart Literature Review on Aquaculture Research	63
Javier De la Hoz-M ⁽¹⁾ , M ^a José Fernández-Gómez ⁽²⁾ , Susana Mendes ⁽³⁾	63



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Análisis multivariante del uso de las TIC y las estrategias de aprendizaje en futuros ingenieros	65
María-Concepción Vega-Hernández ⁽¹⁾ , Carmen Patino-Alonso ⁽²⁾ , Jose-Luis Molina ⁽³⁾ , María-Purificación Galindo-Villardón ⁽⁴⁾	65
Comparación de las distribuciones EBW y Nielsen	67
José María Rodríguez Reinoso ⁽¹⁾ , Julia Rodríguez Reinoso ⁽²⁾ , Valentina Cueva López ⁽³⁾	67
ESTADÍSTICA TEÓRICA	69
Building optimization algorithms for high-dimensional robust linear regression	71
María Jaenada ⁽¹⁾ , Elena Castilla ⁽²⁾	71
Uso del estado del arte de machine learning para hacer inferencia sobre una encuesta del COVID-19	73
Luis Castro Martín ⁽¹⁾ , Ramón Ferri García ⁽²⁾ y María del Mar Rueda García ⁽³⁾	73
Biplot logístico usando algoritmos de machine learning	74
Jose Giovany Babativa-Márquez ⁽¹⁾ , Jose Luis Vicente-Villardón ⁽²⁾	74
CIENCIAS DE LA EDUCACIÓN	77
Determinantes de la competencia matemática de los estudiantes andaluces e implicaciones para el profesorado	79
David Molina-Muñoz ⁽¹⁾ , Antonio Jesús López-Montoya ⁽²⁾ , José Antonio Sánchez-Peigrín ⁽³⁾ , Francisco Martínez-Ortiz ⁽⁴⁾ , Elena Molina-Portillo ⁽⁵⁾	79
La Ciencia de datos en el currículo preuniversitario	80
José Antonio Sánchez Peigrín ⁽¹⁾ , Felipe Ruz ⁽²⁾ , David Molina Muñoz ⁽³⁾ , Antonio Jesús López Montoya ⁽⁴⁾ , José Miguel Contreras García ⁽⁵⁾	80
Análisis de las <i>fake news</i> relacionadas con el Coronavirus desde la perspectiva de las estadísticas cívicas	81
Francisco Martínez-Ortiz ⁽¹⁾ , Felipe Ruz ⁽²⁾ , Antonio Jesús López-Montoya ⁽³⁾ , José Antonio Sánchez Peigrín ⁽⁴⁾ , David Molina-Muñoz ⁽⁵⁾	81



Universidad de Jaén
Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén
Escuela de Doctorado

Saludo del presidente

La ciencia y la tecnología, están ahora más vivas y en movimiento que nunca. El amplio conocimiento de las disciplinas científicas en la actualidad, ha hecho necesaria la especialización en determinados campos, sin embargo, sigue habiendo materias que son tangenciales e imprescindibles en todos los ámbitos científicos. La Estadística es el mejor ejemplo de ello, hay múltiples técnicas estadísticas para el análisis de la información y se ha convertido en una herramienta imprescindible en cualquier publicación de índole científica que se precie. Además, es tan importante que casi no existe actividad humana en que no esté involucrada la Estadística. Las decisiones más importantes de nuestra vida se toman con base en la aplicación de la Estadística.

El trabajo interdisciplinar es fundamental en el proceso de investigación, aúna el conocimiento de distintas materias con un objetivo común; de aquí, es fácil deducir la importancia del intercambio de ideas entre investigadores.

En las tres ediciones anteriores de estas Jornadas (cursos 17-18, 18-19 y 19-20), hemos percibido una buena y positiva acogida en la que se han tratado diversos temas, tales como economía y finanzas, ingeniería, docencia, o ciencias de la salud, cuyo punto en común es el uso indispensable de herramientas estadísticas para obtener resultados. Por tanto, cabe destacar que, la Estadística es una pieza clave en el desarrollo e investigación de ciencia y tecnología y proponemos, con esta segunda edición de las Jornadas, seguir construyendo un foro de discusión e intercambio científico, donde poner en común los diferentes métodos estadísticos que se utilizan o desarrollan en cualquier disciplina científica.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Comité Organizador

Presidente

José Rodríguez Avi (Universidad de Jaén)

Vocales

Valentina Cueva López (Universidad de Jaén)

Irene García Garrido (Universidad de Granada)

Antonio Jesús López Montoya (Universidad de Jaén)

María José Olmo Jiménez (Universidad de Jaén)

Antonio Conde Sánchez (Universidad de Jaén)

Ana María Martínez Rodríguez (Universidad de Jaén)

Beatriz Cobo Rodríguez (Universidad de Granada)

Pedro Antonio Cano Chica (Universidad de Jaén)

Elena Molina Portillo (Universidad de Granada)

David Molina Muñoz (Universidad de Granada)



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Comité Científico

Presidente

José Rodríguez Avi (Universidad de Jaén)

Vocales

Juan Carlos Ruiz Molina (Universidad de Jaén)

Jesús Navarro Moreno (Universidad de Jaén)

Luis Parras Guijosa (Universidad de Jaén)

Emilio Lozano Aguilera (Universidad de Jaén)

Ana Rosa Torres Moreno (Universidad Central de Colombia)

Elias Moreno Bas (Universidad de Granada)

Rafael Herrerías Pleguezuelo (Universidad de Granada)

Juan Manuel Muñoz Pichardo (Universidad de Sevilla)

Antonio Pascual Acosta (Universidad de Sevilla)

Antonio Morillas Márquez (Instituto Nacional de Estadística)

Ana María Aguilera del Pino (Universidad de Granada)

Antonia Oya Lechuga (Universidad de Jaén)

Rosa María Fernández Alcalá (Universidad de Jaén)

María del Mar Rueda García (Universidad de Granada)

María Pilar Frías Bustamante (Universidad de Jaén)



Universidad de Jaén
Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén
Escuela de Doctorado

Programa

MIÉRCOLES 24 DE MARZO

(Enlace: <https://meet.google.com/rso-oqve-pzi>)

09:00-09:15 Inauguración de las IV Jornadas de Estadística como Herramienta Científica.

09:15-10:45 Conferencia Invitada de D. Emilio D. Lozano Aguilera (Director del Departamento de Estadística e Investigación Operativa de la Universidad de Jaén)

“Introducción a la Demografía como herramienta de trabajo”

10:45-11:15 Descanso

11:15-12:45 Conferencia Invitada de D. Antonio J. López Montoya (Profesor del Departamento de Estadística e Investigación Operativa de la Universidad de Jaén)

“Modelización estadística de los fenómenos dependientes de la densidad en poblaciones de animales.”

12:45-14:00 Sesión de ponencias: Ciencias de la Salud

- "Bar-Lev, Bobovitch, and Boukai model to study sexual behavior", Beatriz Cobo et al.
- "Modelos de elección discreta y consumo de tabaco", Imanol L. Nieto González
- "Propuesta de un modelo de salud mental considerando dimensiones psicológicas en tiempos de pandemia", Armando González Sánchez et al.
- "Un algoritmo para estimar los parámetros de un biplot logístico aplicado a la caracterización de distintas variedades de *Colletotrichum graminicola*", Laura Vicente-Gonzalez et al.

14:00-16:00 Almuerzo

16:00-17:45 Sesión de ponencias: Ciencias Económicas

- "Fuzzy Cognitive Maps como herramienta aplicada en la toma de decisiones bajo alta incertidumbre", Julio Vena Oya et al.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA CIENTÍFICA

COMO



Universidad de Jaén

Escuela de Doctorado

- "Consumo de Aceite de Oliva Virgen Extra en Andalucía mediante un Modelo Econométrico ", Manuela Cortés
- "Regresión logística aplicada a las finanzas del fútbol español", Alberto Calahorra López et al.
- "Análisis Estadístico de Datos Textuales sobre Estrategias de Comunicación Empresarial a Nivel Sectorial hacia el Compromiso con la Agenda 2030", Víctor Amor-Esteban et al.
- "Una aproximación estadística de los grandes desequilibrios en el desarrollo socioeconómico del territorio portugués", Miguel Marques de Sousa

17:50-18:30 Sesión de ponencias: Ingeniería

- "Distribuciones Tipo Fase Lineales para la Modelización Probabilística de Datos Funcionales en Electrónica", Christian J. Acal González et al.
- "Modelado estadístico de datos de error posicional. Un ejemplo práctico", Julia Rodríguez-Reinoso et al.

18:30-19:30 Sesión de ponencias: Ciencias Sociales I

- "Análisis Multivariante de las políticas de cierre y contención efectivas frente al COVID-19: países de Europa", Joel Antonio Martínez-Regalado et al.
- "Responsabilidad social ambiental según el Global Reporting Initiative (GRI): un análisis multivariante de las empresas más grandes a nivel mundial", Cinthia L. Murillo-Avalos et al.
- "Técnicas para reducir el sesgo de selección en una encuesta online sobre adicciones", Ramón Ferri-García et al.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

JUEVES 25 DE MARZO

(Enlace: <https://meet.google.com/rso-oqve-pzj>)

09:30-11:00 Conferencia Invitada de Dña. Irene García Garrido (*Profesor del Departamento de Métodos Cuantitativos de la Universidad de Granada*)

“Métodos de estimación en sistemas lineales en tiempo discreto con fallos aleatorios.”

11:00-11:30 Descanso

11:30-12:15 Sesión de ponencias: Ciencias Sociales II

- "Análisis multivariante de la labor de la psicología en los Centros de Reconocimiento de Conductores", Armando González Sánchez
- "Caracterización multivariante del uso de las TIC en estudiantes universitarios de Ciencias Sociales", María-Concepción Vega-Hernández et al.

12:30-14:00 Sesión de póster

- "Caracterización de la desigualdad de género. Una perspectiva multivariante usando HJ-Biplot", Nerea González-García et al.
- "Estudio del Índice Desarrollo Humano a través del Clustering Disjoint HJ-Biplot", Ana B. Nieto-Librero et al.
- "Aplicación de técnicas estadísticas multivariantes y de reducción de dimensionalidad en variables de carga externa y contextuales en un equipo de fútbol semiprofesional", Enrique Benítez Andrés et al.
- "Using a Bayesian Probabilistic Model and Machine Learning Tools for Smart Literature Review on Aquaculture Research", Javier De la Hoz-M et al.
- "Análisis multivariante del uso de las TIC y las estrategias de aprendizaje en futuros ingenieros", María-Concepción Vega-Hernández et al.
- "Comparación de las distribuciones EBW y Nielsen", José María Rodríguez-Reinoso et al.

14:15-16:15 Almuerzo

16:15-17:15 Sesión de ponencias: Estadística Teórica

- "Building optimization algorithms for high-dimensional robust linear regression", María Jaenada y Elena Castilla



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

- "Uso del estado del arte de machine learning para hacer inferencia sobre una encuesta del COVID-19", Luis Castro Martín et al.
- "Biplot logístico usando algoritmos de machine learning", Jose Giovany Babativa-Márquez y Jose Luis Vicente-Villardón

17:30-18:30 Sesión de ponencias: Ciencias de la Educación

- "Determinantes de la competencia matemática de los estudiantes andaluces e implicaciones para el profesorado", David Molina-Muñoz et al.
- "La Ciencia de datos en el currículo preuniversitario", José Antonio Sánchez Pelegrín et al.
- "Análisis de fake news relacionadas con el Coronavirus desde la perspectiva de las estadísticas cívicas", Francisco Martínez-Ortiz et al.



Universidad de Jaén
Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén
Escuela de Doctorado

CIENCIAS DE LA SALUD



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Bar-Lev, Bobovitch, and Boukai model to study sexual behavior

Beatriz Cobo⁽¹⁾, María del Mar Rueda⁽²⁾, Francisca López-Torrecillas⁽³⁾

(1) *Departamento de Métodos Cuantitativos para la Economía y la Empresa, Universidad de Granada, e-mail: beacr@ugr.es*

(2) *Departamento de Estadística e Investigación Operativa, Universidad de Granada, e-mail: mrueda@ugr.es*

(3) *Departamento de Personalidad, Evaluación y Tratamiento Psicológico, Universidad de Granada, e-mail: fcalopez@ugr.es*

RESUMEN

El comportamiento sexual es difícil de estudiar empíricamente debido a su naturaleza sensible. La prevalencia y la frecuencia del comportamiento sexual son difíciles de estimar mediante técnicas de encuesta estándar porque los encuestados tienden a ocultar información en tales entornos.

El sesgo de deseabilidad social (el deseo de causar una impresión favorable) representa una amenaza significativa para la validez de los informes, particularmente cuando se refieren a información sensible relacionada con actividades sexuales.

Desde la década de 1960, se han ideado una variedad de métodos de interrogatorio para garantizar el anonimato de los encuestados y reducir la incidencia de respuestas evasivas y la denuncia excesiva o insuficiente de actos socialmente indeseables. Estos métodos, generalmente conocidos como técnicas de interrogatorio indirecto, obedecen al principio de que no se plantea ninguna pregunta directa a los participantes de la encuesta. Por lo tanto, no es necesario que los encuestados revelen abiertamente si realmente se han involucrado en actividades o actitudes que son socialmente sensibles.

En el presente estudio, se utilizan métodos de interrogatorio directo e indirecto, específicamente el modelo de Bar-Lev, Bobovitch y Boukai, para explorar informes de comportamiento sexual inapropiado por parte de estudiantes universitarios. Las estimaciones indirectas del número de veces que los estudiantes no pudieron contener su comportamiento sexual inapropiado fueron significativamente más altas que las estimaciones directas, y también entre estudiantes masculinos como femeninos. Los resultados obtenidos sugieren que el método indirecto genera valores más altos de informes



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA COMO HERRAMIENTA CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

autoestigmatizantes de experiencias sexuales al aumentar la privacidad en el proceso de recolección de datos, por lo que queda demostrado que las técnicas indirectas son un método útil para investigar el comportamiento sexual.

Palabras clave: comportamiento sexual, sesgo de deseabilidad social, técnicas de interrogatorio indirecto, modelo de Bar-Lev, Bobovitch, and Boukai.

ABSTRACT

Sexual behavior is difficult to study empirically because of its sensitive nature. The prevalence and frequency of sexual behavior are difficult to estimate by standard survey techniques because respondents tend to withhold information in such settings.

Social desirability bias (the wish to make a favorable impression) poses a significant threat to the validity of self-reports, particularly when they concern sensitive information related to sexual activities.

Since the 1960s, a variety of questioning methods have been devised to ensure respondents' anonymity and to reduce the incidence of evasive answers and the over/ underreporting of socially undesirable acts. These methods, generally known as indirect questioning techniques, obey the principle that no direct question is posed to survey participants. Therefore, there is no need for respondents to openly reveal whether they have actually engaged in activities or present attitudes that are socially sensitive.

In the present study, direct and indirect questioning methods, specifically Bar-Lev, Bobovitch, and Boukai model, are used to explore reports of inappropriate sexual behavior by university students. The indirect estimates of the number of times that the students were unable to restrain their inappropriate sexual behavior were significantly higher than the direct estimates, and also among both male and female students. The results obtained suggest that the indirect method elicits higher values of self-stigmatizing reports of sexual experiences by increasing privacy in the data collection process, which is why it is shown that indirect techniques are a useful method to investigate sexual behavior.

Keywords: sexual behavior, social desirability bias, indirect questioning techniques, Bar-Lev, Bobovitch, and Boukai model.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA CIENTÍFICA

COMO



Universidad de Jaén

Escuela de Doctorado

Modelos de elección discreta y consumo de tabaco

Imanol L. Nieto González

*Facultad de Economía, Empresa y Turismo, Universidad de La Laguna, alu0100801811@ull.edu.es –
inietoglez@gmail.com*

RESUMEN

El tabaquismo supone un problema de salud pública a nivel mundial, por ello, su solución pasa por la elaboración de estrategias de prevención fundamentadas en evidencia empírica y orientadas a colectivos concretos. Los modelos de elección discreta son una metodología apropiada para el análisis de la decisión que toman los fumadores, en base a su probabilidad de elección, como el fundamento sobre el que se sustente la propuesta de estos planes preventivos. A partir de la Encuesta Nacional de Salud (INE) y de la estimación de modelos logit multinomiales se contrasta que variables de tipo sociodemográfico, como el género o nivel de estudios, resultan claves a la hora de orientar adecuadamente estas estrategias.

Palabras clave: modelos de elección discreta, microeconometría, tabaco.

ABSTRACT

Smoking is a global public health problem, so its solution requires the development of prevention strategies based on empirical evidence and targeted to specific groups. Discrete choice models are a proper methodology for the analysis of the decision of smokers, based on their probability of choice, as the basis for the proposal of these preventive plans. Based on the National Health Survey (INE) and the estimation of multinomial logit models, it is shown that sociodemographic variables, such as gender or level of education, which are key when it comes to properly orienting these strategies.

Keywords: discrete choice models, microeconomic, tobacco.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA CIENTÍFICA

COMO



Universidad de Jaén

Escuela de Doctorado

Propuesta de un modelo de salud mental considerando dimensiones psicológicas en tiempos de pandemia.

Armando González Sánchez^{† (1)}, Greibin Villegas Barahona⁽²⁾ M.Sc. Eva Carazo Vargas,⁽³⁾ Máster Raúl Ortega⁽⁴⁾, Harold Arias LeClaire⁽⁵⁾.

(1) Departamento de estadística, Universidad de Salamanca, armando_gonzalez@usal.es

(2) Departamento de estadística, Universidad Estatal a Distancia de Costa Rica

(3) Vicerrectoría de Investigación, Universidad Estatal a Distancia de Costa Rica

(4) Universidad Nacional de Costa Rica

(5) Escuela de Ciencias Exactas y Naturales, Universidad Estatal a Distancia de Costa Rica

RESUMEN

Con el uso de las escalas psicométricas GAD-2, PHQ-2, de miedo social a la COVID-19, así como adherencia a las medidas de protección, se han utilizado para conformar grupos de individuos en función del nivel de afectación en la salud mental. Se han conformado 4 grupos de afectación según sus características particulares. Mediante un análisis multivariante se seleccionaron los 13 ítems más relevantes desde el punto de vista de la clasificación de la salud mental para construir un modelo de clasificación con niveles superiores al 90% de clasificación correcta. En este modelo interesa que el grupo de mayor afectación en salud mental tenga porcentajes de buena discriminación.

Palabras clave: salud mental, COVID, Clúster, multivariante

ABSTRACT

With the use of the psychometric scales GAD-2, PHQ-2, Social Fear of COVID-19, plus Adherence to Protection Arrangements, they have been used to form groups of individuals based on the level of mental health affectation. 4 groups of affectation have been formed according to their particular characteristics. Through a multivariate analysis, the 13 most relevant items from the mental health classification point of view were selected to construct a classification model with levels higher than 90% of correct classification. In this model, it is interesting that the group with the greatest impact on mental health has percentages of good discrimination.

Keywords: mental health, COVID, cluster, multivariant



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Un algoritmo para estimar los parámetros de un biplot logístico
aplicado a la caracterización de distintas variedades de *Colletotrichum
graminicola*

Laura Vicente-Gonzalez⁽¹⁾, Michael R. Thon⁽²⁾, Serenella A. Sukno⁽³⁾ & José L. Vicente-Villardón⁽⁴⁾

(1) Departamento de Estadística, Universidad de Salamanca, e-mail: laura20vg@usal.es

(2) CIALE, Universidad de Salamanca, e-mail: mthon@usal.es

(3) CIALE, Universidad de Salamanca, e-mail: ssukno@usal.es

(4) Departamento de Estadística, Universidad de Salamanca, e-mail: villardon@usal.es

RESUMEN

Es cada vez más frecuente encontrar grandes matrices de datos con un número elevado de variables, incluso mayor que el número de individuos. Cuando se trata de establecer diferencias entre grupos deberían utilizarse métodos multivariantes que son los más adecuados para dar una imagen general de los datos y controlar el Riesgo Tipo I. El método más utilizado para realizar este tipo de análisis es el Análisis Multivariante de la Varianza (MANOVA) basado en el Modelo Lineal General Multivariante (MGLM) que es de aplicación limitada para este tipo de datos y solamente es útil cuando tenemos variables cuantitativas, sin embargo, existen un gran número de técnicas alternativas que pueden ser empleadas cuando las condiciones básicas de aplicación del MANOVA no se verifican.

En este trabajo se van a aplicar técnicas alternativas al MANOVA desarrolladas en los últimos años para el estudio de un conjunto de datos de secuenciación mediante la técnica RAD-Seq (Restriction Associated DNA Sequencing) de *Colletotrichum graminicola*.

Se trata de un hongo que puede hospedarse en un gran número de especies diferentes, aunque el hospedador principal es el maíz donde provoca una enfermedad denominada antracnosis.

Para el análisis disponemos de una matriz de datos binarios compuesta por 103 filas (variedades del hongo) y 13183 columnas (polimorfismos genéticos obtenidos de la secuenciación). Las 103 filas están clasificadas en función de su origen (9 países o 3 continentes). Los objetivos son: estudiar si las variedades que tienen el mismo origen son más similares que aquellas variedades que tienen diferentes orígenes, y qué variedades y orígenes son más similares.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA COMO HERRAMIENTA CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Para establecer la significación de la diferencia entre grupos, se usará PERMANOVA (Anderson, 2001, 2005) y/o BOOTMANOVA, análisis multivariantes basados en distancias y representaciones gráficas asociadas a los mismos.

Además de contrastes de asociación para cada variable por separado, las técnicas más utilizadas en estudios de este tipo son el Análisis de Cluster (CA) y en menor medida técnicas de ordenación como el Análisis de Coordenadas Principales (PCoA). Ambos métodos se centran en la clasificación de los individuos, pero no determinan que variables son responsables de la misma.

Una variación de PCoA, denominada Biplot Logístico Externo (Demey et al., 2008), incluye información sobre las variables junto con la clasificación de los genotipos. En este método, las coordenadas de los individuos y las variables se calculan para tener respuestas logísticas a lo largo de las dimensiones del PCoA. El método está relacionado con la regresión logística de la misma manera que el Análisis Biplot Clásico (CBA) está relacionado con la regresión lineal, por lo que el método se ha denominado "Biplot logístico" (LB), el método es externo porque las coordenadas de filas y columnas se calculan en procedimientos separados.

En este trabajo proponemos un algoritmo para estimar simultáneamente las coordenadas de filas y columnas basado en los métodos de descenso de gradiente que generaliza la propuesta de Vicente-Villardón et al. (2006) y obtiene una bondad de ajuste mejor que los métodos externos.

Los análisis se realizarán con el Software R (R Core Team, 2016), empleando paquetes desarrollados por el equipo (Vicente-Gonzalez & Vicente-Villardón, 2021; Vicente-Villardón, 2021).

Con las técnicas propuestas conseguimos la caracterización de las variedades y mostrar las diferencias entre las procedencias.

Palabras clave: PERMANOVA, BOOTMANOVA, Logistic Biplot, PCoA, Biplot, Bootstrap, Distancias, *Colletotrichum graminicola*



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA CIENTÍFICA

COMO



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

Due to recent advances in data collection, it is every time more frequent to have data matrices with a high number of variables, even higher than the number of individuals. When the aim of the study is to establish the significance of the differences among several groups, multivariate analysis should be used to give a general picture of the data and to control the Type I risk. The most widely used method to perform this type of analysis is the Multivariate Analysis of Variance (MANOVA) based on the Multivariate General Linear Model (MGLM) that has been developed for numerical data and has limited applicability for big data matrices. Nevertheless, there are a large number of alternative techniques that can be used in these cases.

In this work, alternative techniques to MANOVA, developed in recent years, will be applied to study a set of data sequenced using the RAD-Seq technique (Restriction Associated DNA Sequencing) of *Colletotrichum graminicola*. It is a fungus that a large number of species can host, being maize one of the most important, where it causes a disease called anthracnose.

For the analysis we have a binary data matrix made up of 103 rows (varieties) and 13183 columns (genetic polymorphisms). The 103 rows are classified according to their origin. The proposed objectives are to study if varieties of the same origin are more similar than varieties of different origins, and what varieties and origins are more similar.

To establish the significance of the differences among groups, we have used PERMANOVA (Anderson, 2001, 2005) and/or BOOTMANOVA, multivariate analysis based on distances.

Together with separate contrasts for each variable, the most used techniques in previous studies are Cluster Analysis (CA) and Principal Coordinate Analysis (PCoA). Both methods focus on the classification of individuals, but do not have information on which variables are responsible for the classification.

A variation of PCoA, called External Logistic Biplot (Demey et al., 2008), includes information about the variables along with the classification of the genotypes. The coordinates of individuals and variables are computed to have logistic responses along the biplot dimensions. The method is related to logistic regression in the same way that Classical Biplot Analysis (CBA) is related to linear regression. Thus, the method was named "External Logistic Biplot" (LB). The method is external because coordinates for rows and columns are calculated in separate procedures.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA CIENTÍFICA

COMO



Universidad de Jaén

Escuela de Doctorado

In this paper, we propose an algorithm to simultaneously estimate row and column coordinates based on gradient descent methods, generalizing the procedure proposed by (Vicente-Villardón et al., 2006).

The analysis is done in R (R Core Team, 2016), with packages developed by the team (Vicente-Gonzalez & Vicente-Villardón, 2021; Vicente-Villardón, 2021)

With the proposed techniques the differences among countries have been established.

Keywords: PERMANOVA, Logistic Biplot, PCoA, Biplot, Bootstrap, Distances, *Colletotrichum graminicola*

BIBLIOGRAFÍA

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral ecology*, 26(1), 32-46.

Anderson, M. J. (2005). Permutational multivariate analysis of variance. *Department of Statistics, University of Auckland, Auckland*, 26, 32-46.

Demey, J. R., Vicente-Villardón, J. L., Galindo-Villardón, M. P., & Zambrano, A. Y. (2008). Identifying molecular markers associated with classification of genotypes by External Logistic Biplots. *Bioinformatics*, 24(24), 2832-2838.
<https://doi.org/10.1093/bioinformatics/btn552>

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>

Vicente-Gonzalez, L., & Vicente-Villardón, J. L. (2021). *PERMANOVA: Multivariate Analysis of Variance Based on Distances and Permutations*. <https://CRAN.R-project.org/package=PERMANOVA>

Vicente-Villardón, J. L. (2021). *MultBiplotR: MULTivariate Analysis Using BIPLoTs*.

Vicente-Villardón, J. L., Galindo-Villardón, M. P., & Blázquez-Zaballos, A. (2006). Logistic Biplots. *Múltiple Correspondence Análisis And Related Methods*.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

CIENCIAS ECONÓMICAS



Fuzzy Cognitive Maps como herramienta aplicada en la toma de decisiones bajo alta incertidumbre

Julio Vena Oya¹⁾, José Alberto Castañeda García²⁾, Miguel Ángel Rodríguez Molina³⁾

(1) Organización de empresas, marketing y sociología, Universidad de Jaén, e-mail: jvena@ujaen.es

(2) Departamento de comercialización e investigación de mercados, Universidad de Granada, e-mail: jalberto@ugr.es

(3) Departamento de comercialización e investigación de mercados, Universidad de Granada, e-mail: rmolina@ugr.es

RESUMEN

Los gestores y políticos encargados de tomar decisiones y liderar en situaciones complejas a menudo se encuentran ante situaciones de alta incertidumbre donde la información de la que disponen es limitada o nula, llevando la pandemia de COVID-19 esta circunstancia a un nuevo nivel, con niveles de incertidumbre no vistos anteriormente, en un contexto del que no tenemos información previa en el que la toma de decisiones a menudo conlleva un alto coste.

A pesar de que los modelos econométricos están siendo empleados para predecir escenarios relacionados con las consecuencias de la pandemia, éstos no tienen en cuenta las medidas que semanalmente están tomando políticos y gestores. Por ello, este trabajo trata de aplicar un enfoque semi-cuantitativo *fuzzy* para desarrollar un método que provea de escenarios cuyas previsiones sean consistentes y creíbles a largo plazo para facilitar la toma de decisiones bajo elevada incertidumbre. El contexto escogido para desarrollar esta técnica será el del turismo internacional español y las consecuencias de la COVID-19 en el mismo.

A través de la técnica de *Fuzzy Cognitive Maps* se elaboran 3 escenarios empíricos robustos que tratan de predecir la probabilidad de una crisis económica más severa que la de 2008 como consecuencia de la pandemia.

Esta técnica puede ayudar a los investigadores a implementar sistemas de predicción que traten de vislumbrar el impacto potencial de eventos difusos o inciertos a través de la construcción de escenarios flexibles con alta capacidad de adaptación a ambientes inestables.

Palabras clave: Fuzzy Cognitive Maps, métodos semi-cuantitativos, gestión de crisis, turismo internacional



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

Managers and policymakers are usually facing to uncertain and fuzzy situations where they must to make decisions under unprecedented context or scarce with imperfect information. Nevertheless, COVID-19 pandemic has put the level of complexity in a higher level, dealing with an environment with no previous information and where the cost of making mistakes is too high.

Econometric models have been largely used to face with this pandemic situations, trying to forecast some scenarios but omitting the weekly information weekly provided by experts and policymakers. So, this research aim is to apply a fuzzy semi-quantitate method approach to develop a method that try to fix this uncertain situation, choosing the context of Spanish inbound tourism and the consequences of the pandemic in this sector.

Using fuzzy cognitive mapping, a model with three robust empirical scenarios is constructing to forecast the probability of an economic crisis in Spain, more severe than the 2008 GFC. This semi-quantitative model can help researchers to forecast the potential impact of major events in fuzzy or uncertain environments by constructing flexible scenarios with the capacity to adapt to unstable scenarios.

Keywords: Fuzzy Cognitive Maps, semi-quantitative method, crisis management, international tourism



Consumo de Aceite de Oliva Virgen Extra en Andalucía mediante un Modelo Econométrico

Manuela Cortés⁽¹⁾

(1) Departamento de Estadística e Investigación Operativa, Universidad de Jaén, info@ujaen.es

RESUMEN

La realización de este trabajo consiste en estudiar la evolución acerca del Consumo del Aceite de Oliva Virgen Extra en Andalucía, mediante un modelo econométrico, el cual está formado por una serie de variables macroeconómicas como son: Precio Medio, Penetración e Índice de Precios de Consumo (IPC).

Destacando principalmente nuestra Comunidad Autónoma líder en el Consumo de Aceite de Aceite de Oliva Virgen Extra seguido de Madrid y Extremadura.

En primer lugar, se deben recopilar los datos necesarios para poder realizar nuestro estudio y partir de ahí poder construir el modelo econométrico, enfocándolo al periodo de tiempo del estudio, el cual comprende desde Enero 2010 hasta Diciembre de 2019.

En el modelo podrán verse reflejadas todas las variables que hemos elegido, aunque todas no tienen que ser útiles y decidiremos si permanece en el modelo o por el contrario las eliminaremos hasta conseguir determinar un modelo que describa de manera próxima a la realidad, el cual cumpla todas las hipótesis básicas, mediante la aplicación de las diferentes fases del método econométrico. Finalmente, se elaboran las conclusiones en las cuales son detallados e interpretados los resultados obtenidos en el análisis.

Palabras clave: Consumo, Aceite de Oliva Virgen Extra, Modelo Econométrico, Comunidad Autónoma.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

The realization of this work consists of studying the evolution regarding the Consumption of Extra Virgin Olive Oil in Andalusia, through an econometric model, which is formed by a series of macroeconomic variables such as: Average Price, Penetration and Consumer Price Index (CPI). Mainly highlighting our Autonomous Community leader in the Consumption of Extra Virgin Olive Oil followed by Madrid and Extremadura. In the first place, the necessary data must be collected to be able to carry out our study and from there be able to build the econometric model focusing in on the period of time of the study, which ranges from January 2010 to December 2019.

In the model, all the variables that we have chosen can be reflected, although they do not all have to be useful and we will decide if it remains in the model or on the contrary we will eliminate them until we are able to determinate a model that describes closely to reality, which fulfil all the basic hypotheses, by applying the different phases of the econometric method.

Finally, the conclusions are drawn up in which the results obtained in the analysis are detailed and interpreted.

Keywords: Consumption, Extra Virgin Olive Oil, Econometric Model, Autonomous Community.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA CIENTÍFICA

COMO



Universidad de Jaén

Escuela de Doctorado

Regresión logística aplicada a las finanzas del fútbol español

Alberto Calahorro López (1), Julio Vena Oya (2), Melinda Ratkai, PhD (3)

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

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

(3) Les Roches Education, Marbella, e-mail: Melinda.ratkai@lesroches.es

RESUMEN

El propósito de este estudio es analizar si las regulaciones de 'Juego Limpio Financiero' (FFP) establecidas por la UEFA 2011, podrían haber afectado a las finanzas de los clubes de fútbol. Para ello, a través de una regresión logística se observa si la probabilidad de participar en competiciones europeas puede explicarse a través de la situación financiera de los mismos, comprobando si existen diferencias antes y después de la aplicación de esta regulación. Los resultados muestran diferencias en la estructura financiera de los clubes analizados y su relación con la probabilidad de clasificar para Europa antes y después del FFP. Con ello, se trata de aportar evidencia sobre los efectos del 'Juego Limpio Financiero' en los clubes de fútbol de la liga española.

Palabras clave: juego limpio financiero; regresión logística; finanzas; Liga Española, competiciones europeas

ABSTRACT

The purpose of this study is to analyze whether the Financial Fair Play (FFP) regulations established by UEFA 2011 may have affected the finances of soccer clubs. For this purpose, through a logistic regression it was observed whether the probability of participating in European competitions might be explained by the financial situation of the clubs, checking if there are differences before and after the application of this regulation. The results show differences in the financial structure of the clubs analyzed and their relationship with the probability of qualifying for Europe before and after FFP. The aim is to provide evidence on the effects of 'Financial Fair Play' on Spanish league soccer clubs.

Keywords: Financial Fair Play, logistic regression, finance, Spanish League, European competitions.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA COMO HERRAMIENTA CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

INTRODUCCIÓN

Durante los años 80 y 90, los clubes de fútbol han estado atravesando dificultades financieras, agravado por la sentencia Bosman de 1996 (Solberg y Haugen, 2010). Según Barajas y Rodríguez (2010), la crisis financiera coincidió en el tiempo con aumentos en los ingresos, el problema reside en el desequilibrio entre ingresos y gastos y, como consecuencia, el aumento de la deuda.

La crisis financiera generalizada ha implicado dificultades crecientes para hacer frente a la financiación. Tras varios intentos fallidos de reducir los problemas financieros que atravesaban los equipos de las grandes ligas por parte de las administraciones públicas, la entrada del FFP ha conseguido sanear la salud financiera de los clubes europeos (Franck, 2014). Sin embargo, existen autores que denuncian que ha aumentado la diferencia entre clubes que consiguen participar en competiciones europeas y aquéllos que no (Birkhäuser, Kaserer y Urban, 2019; Dimitropoulos y Scafarto, 2021). Por todo ello, nuestro objetivo es comprobar si la probabilidad de clasificarse para estas competiciones europeas puede explicarse a través de la situación financiera de los clubes, concretamente a través de las ratios de endeudamiento, liquidez y solvencia durante las temporadas 2005-2019 de la liga española.

METODOLOGÍA

A través de la base de datos SABI se han obtenido los siguientes datos financieros: ratio de endeudamiento, liquidez y solvencia, (Beech, Horsman, y Magraw, 2010; Szymanski, 2017) como indicadores de salud financiera en base a las cuentas anuales de los clubes que han participado en la primera división de la liga española durante el periodo 2005-2019.

A través de una regresión logística se trata de determinar la relación entre estos ratios y la probabilidad de participar en competiciones europeas. Esta técnica es ampliamente utilizada artículos de finanzas.

A través del software SPSS-25 se formula un modelo de regresión logística (Landau y Everitt, 2017, p221) siendo la variable dependiente la clasificación (CE=1) o no (CE=0) para competición europea, siendo el modelo a ajustar:

$$y = \pi(x) + \epsilon$$



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA CIENTÍFICA

COMO



Universidad de Jaén

Escuela de Doctorado

Donde $\pi(x)$ es la probabilidad de clasificarse para CE en función de una variable independiente x_i , es decir: $P(Y=1/x)$. Este tipo de modelo establece que la función de regresión $\pi(x)$ sigue la siguiente forma:

$$\pi(x) = \frac{1}{1 + (e^{-(\beta_0 + \beta_1 x)})} = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$$

RESULTADOS:

Se observa cómo antes de la aplicación del FFP, los equipos que disponían de capacidad de endeudamiento, solvencia (para poder endeudarse) y líquido (consecuencia de la entrada de dinero por el mayor endeudamiento) tenían mayores probabilidades de clasificarse para Europa, por el contrario, con la entrada en vigor de esta norma, se observa como los ratios de endeudamiento y liquidez ya no influyen y son los equipos que disponen de líquido los que mayor probabilidad de clasificación presentan. Esto puede deberse a...

Referencias:

Barajas, Á., y Rodríguez, P. (2010). Spanish football clubs' finances: Crisis and player salaries. *International Journal of Sport Finance*, 5(1), 52.

Beech, J., Horsman, S., y Magraw, J. (2010). Insolvency events among English football clubs. *International Journal of Sports Marketing and Sponsorship*.

Birkhäuser, S., Kaserer, C., y Urban, D. (2019). Did UEFA's financial fair play harm competition in European football leagues?. *Review of Managerial Science*, 13(1), 113-145.

Dimitropoulos, P., y Scafarto, V. (2021). The impact of UEFA financial fair play on player expenditures, sporting success and financial performance: evidence from the Italian top league. *European Sport Management Quarterly*, 21(1), 20-38.

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

Landau, S. y Everitt, B.S. (2017). A handbook of statistical analyses using SPSS. Chapman y Hall.

Solberg, H. A., y Haugen, K. K. (2010). European club football: Why enormous revenues are not enough?. *Sport in Society*, 13(2), 329-343.

Szymanski, S. (2017). Entry into exit: insolvency in English professional football. *Scottish Journal of Political Economy*, 64(4), 419-444.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Análisis Estadístico de Datos Textuales sobre Estrategias de Comunicación Empresarial a Nivel Sectorial hacia el Compromiso con la Agenda 2030

Víctor Amor-Esteban ⁽¹⁾, Isabel-María García-Sánchez ⁽²⁾, David Galindo-Álvarez ⁽³⁾

(1) Departamento de Estadística, Universidad de Salamanca, vamor@usal.es

(2) Instituto Multidisciplinar de Empresa, Universidad de Salamanca, lajefa@usal.es

(3) Facultad de Economía y Empresa, Universidad de Salamanca, davidgalindo97@usal.es

RESUMEN

En los últimos años se observa un notable aumento en la inclusión de los Objetivos de Desarrollo Sostenible (ODS) dentro de las estrategias de Responsabilidad Social Corporativa (RSC). Esta investigación propone un análisis estadístico de datos textuales de la comunicación empresarial con sus grupos de interés sobre la contribución a la Agenda 2030 de las 160 empresas que actualmente cotizan en España, con el fin de identificar la posible existencia de prácticas de gestión de impresiones a través de la connotación de sus informes. Este análisis se realiza a nivel sectorial y consiste en identificar los términos comúnmente empleados por las empresas, los datos obtenidos se someten a un proceso de limpieza donde son eliminados los términos vacíos y se aplica una lematización a los elementos textuales semánticamente cargados; esta matriz léxica es representada de manera bidimensional a través de un Análisis Factorial de Correspondencias (AFC), el cual trabaja con perfiles y utiliza la distancia ji-cuadrado, neutralizando todas las distorsiones en la representación gráfica, ya que es una distancia euclídea ponderada y permite que los términos más frecuentes se ponderen menos y los poco frecuentes se ponderen más. Los resultados muestran que empresas con objetivos legales hacen mayor uso de estrategias de gestión de impresiones defensivas desviando la atención de sus grupos de interés con un discurso con connotación negativa, centrado en justificaciones, disculpas o excusas; por el contrario, encontramos informes con connotación positiva en compañías encaminadas a objetivos comerciales y altruistas, basados en la promoción organizativa, la autopromoción y la felicitación.

Palabras clave: Agenda 2030, Objetivos Desarrollo Sostenible (ODS), minería texto, análisis correspondencias, teoría gestión impresiones.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA COMO HERRAMIENTA CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

In recent years, there has been a notable increase in the inclusion of the Sustainable Development Goals (SDGs) within the Corporate Social Responsibility (CSR) strategies. This research proposes a statistical analysis of textual data of business communication with its stakeholders on the contribution to the 2030 Agenda of the 160 companies that are currently listed in Spain, in order to identify the possible existence of impression management practices to through the connotation of your reports. This analysis is carried out at the sectoral level and consists of recognizing the terms commonly used by companies. The data obtained is subjected to a cleaning process where empty terms are eliminated and a lemmatization is applied to the semantically loaded textual elements; this lexical matrix is represented in a two-dimensional way through a Factorial Correspondence Analysis (FCA), which works with profiles and uses the chi-square distance, neutralizing all the distortions in the graphic representation, since it is a weighted Euclidean distance and allows the most frequent terms to be weighted less and the infrequent ones to be weighted more. The results show that companies with legal objectives make greater use of defensive impression management strategies, diverting the attention of their interest groups with a discourse with a negative connotation, focused on justifications, apologies or excuses; on the contrary, we find reports with a positive connotation in companies aimed at commercial and altruistic objectives, based on organizational promotion, self-promotion and congratulations.

Keywords: 2030 Agenda, Sustainable Development Objectives (SDGs), text mining, correspondence analysis, impression management theory.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA COMO HERRAMIENTA CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Una aproximación estadística de los grandes desequilibrios en el desarrollo socioeconómico del territorio portugués

Miguel Marques de Sousa

Departamento de Estadística (doctorando), Universidad de Salamanca, e-mail: id00750038@usal.es

RESUMEN

El que nos mueve para este trabajo es caracterizar las dinámicas que caván profundas asimetrías en el desarrollo económico y social, de diversas partes del espacio continental portugués.

La dinámica de los espacios regionales en el territorio nacional, del mismo modo como las formas organizativas que han adquirido esos territorios, no surgieran por mera casualidad. Durante siglos, factores naturales, económicos, históricos, políticos, étnicos, antropológicos y otros, generaron identidades sociales, políticas, culturales, arquitectónicas, etc. Estos han sido borrosos por procesos más o menos abruptos y, por consecuencia, han seguido dinámicas que han conducido muchos, en la actualidad, para niveles del desarrollo económico y social que ponen en causa su sostenibilidad.

El conjunto de municipios es descrito por un conjunto de 20 variables, con datos del último censo del Instituto Nacional de Estadística de Portugal. Para este análisis estadístico multivariante aplicamos las técnicas del análisis de componentes principales (ACP), con el *software* R. Las variables cubren las áreas de Economía, Demografía, Empresa, Social, Cultural y Cívica. Entre ellas hay cualitativas y cuantitativas suplementarias.

Deseamos aprehender la variabilidad de los municipios de una región a otra de manera multidimensional, para llegar a una noción pertinente de perfil para considerar un municipio. Cuando hablamos de municipios, nos referimos al nivel más bajo de la escala de armonización de unidades territoriales con fines estadísticos de la Unión Europea,



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

denominado "Núcleos de Unidades Territoriales", abreviado como "NUTS". En este caso, se trata del nivel NUT III. Podríamos así responder a cuestiones del tipo: ¿Cuáles son las desigualdades más grandes entre municipios?

¿Podemos construir indicadores sintéticos que permiten caracterizar y resumir las diferencias de perfiles de municipios?

Estos factores podrán servir de base en la construcción de una tipología sobre los municipios.

En el estudio de variables, cada variable mide los niveles de 278 municipios. La relación entre las variables es aprehendida a partir de los niveles de cada una en los municipios y no del conjunto. Naturalmente, queremos obtener una visión del conjunto de estas relaciones, sin analizar cada pareja de variables. Se pretende investigar las principales dimensiones latentes en este conjunto de variables, para permitir la caracterización y agrupación según estas dimensiones permitiendo, a su vez, apoyar el debate sobre las políticas regionales.

Palabras territorio; municipios; NUT; variables; análisis; componentes; PCA; desarrollo; sostenibilidad.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

INGENIERÍA



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Distribuciones Tipo Fase Lineales para la Modelización Probabilística de Datos Funcionales en Electrónica

Christian J. Acal González⁽¹⁾, Juan Eloy Ruiz Castro⁽¹⁾, Ana M. Aguilera del Pino⁽¹⁾

(1) Departamento de Estadística e Investigación Operativa, Universidad de Granada: chracal@ugr.es, jeloy@ugr.es, aaguiler@ugr.es

RESUMEN

En la actualidad, las memorias RRAM aparecen en muchos dispositivos (teléfonos móviles, tablets, portátiles, etc.) que son utilizados en una infinidad de ocasiones a lo largo del día. El número tan elevado de ventas de estos productos ha provocado que los beneficios asociados a las memorias RRAM no paren de aumentar año tras año en este sector, hasta tal punto de convertirse en la principal fuente de ingresos a nivel mundial en la industria de los semiconductores. Sin embargo, existe una gran necesidad de estudiar la variabilidad relacionada con el funcionamiento de estas memorias antes de proceder a los procesos de fabricación y venta de las mismas

. En la práctica, esta variabilidad es traducida en diferentes curvas de voltaje-intensidad. Debido a que, por tanto, los datos experimentales son curvas, una herramienta apropiada para tratar con este tipo de datos es el Análisis de Datos Funcionales (FDA). El FDA engloba una gran variedad de métodos estadísticos que han sido obtenidos como la generalización natural de las técnicas estadísticas multivariantes tradicionales al caso de estar trabajando con curvas en lugar de con vectores. En esta rama, una de las técnicas más importantes debido a la gran dimensión de los datos es el Análisis de Componentes Principales Funcional (FPCA).

Esta técnica tiene como misión reducir la dimensión del problema e interpretar los principales modos de variación de las curvas en términos de un pequeño conjunto de variables incorreladas llamadas componentes principales. Se ha comprobado que las curvas de voltaje-intensidad pueden ser reconstruidas con bastante precisión a través de un número bajo de estas componentes principales. Este resultado es realmente notable y esperanzador desde el punto de vista de la simulación de circuitos para memorias RRAM. Sin embargo, es necesario identificar la distribución que mejor se ajuste a las componentes principales para poder simular tantas curvas como se desee.

A este respecto, en este trabajo se introduce y desarrolla una nueva clase de distribuciones denominada distribuciones Tipo Fase Lineal cuyas propiedades algorítmico-matriciales son estudiadas en profundidad. Esta clase de distribuciones se construye desde las



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

distribuciones Tipo Fase teniendo en este nuevo caso soporte en cualquier semirrecta real. Las distribuciones Tipo Fase Lineales permitirán modelizar las componentes principales y como consecuencia, caracterizar las distribuciones unidimensionales del proceso que genera los datos.

Palabras clave: datos funcionales, componentes principales, distribuciones Tipo Fase, distribuciones Tipo Fase Lineales, memorias RRAM.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Modelado estadístico de datos de error posicional. Un ejemplo práctico

Julia Rodríguez-Reinoso⁽¹⁾, José María Rodríguez-Reinoso⁽²⁾

(1) Foqum analytics. e-mail: juliarodriguezreinoso@gmail.com

(2) Universidad de Granada, e-mail: jose.maria.rodriguez.reinoso@gmail.com

RESUMEN

La hipótesis de normalidad en el caso de medidas de error aparece desde el mismo origen de la propia distribución normal ya que Gauss llegó a ella analizando errores de medición en observaciones astronómicas. El hecho de que unos errores o residuos se distribuyan normalmente implica que éstos se deben al puro azar, y no hay otras causas que los expliquen. Además, la hipótesis de normalidad es básica a la hora de plantear los contrastes de hipótesis sobre errores, tanto para valores medios como para varianzas.

Sin embargo, en la práctica es difícil encontrar datos de errores de medida que se distribuyan según una distribución normal. Una causa es que realmente esos datos procedan de otra distribución, pero en muchos casos se debe a que el conjunto de las observaciones se ha obtenido de diferentes distribuciones normales, con medias y/o varianzas distintas, y la mezcla de todos esos datos en un único conjunto de datos no tiene por qué seguir una única distribución normal.

En estos casos se puede aplicar la técnica de mixtura finita de distribuciones gaussianas. Consiste en descomponer los datos en múltiples distribuciones normales, estimando la media y la varianza correspondiente, así como la probabilidad en la mixtura. De esta manera se puede generar un modelo poblacional que permita conocer mejor la naturaleza de la variable analizada. En este trabajo proponemos una aplicación de este procedimiento para unos datos procedente de un vuelo de campaña ALS en el que se midieron errores sobre diferentes tipos de terreno.

Palabras clave: Distribución normal, Mixtura finita de distribuciones. Estimación EM.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
CIENTÍFICA
COMO



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

The hypothesis of normality in the case of error measurements appears from the beginning of the normal distribution itself, since Gauss arrived at it by analysing measurement errors in astronomical observations. The fact that some errors or residuals are normally distributed implies that they are due to pure chance, and there are no other causes that explain them. In addition, the normality hypothesis is basic when it comes to proposing the contrasts of hypotheses on errors, both for mean values and for variances.

However, in practice it is difficult to find measurement error data that are distributed according to a normal distribution. One cause is that these data really come from another distribution, but in many cases it is due to the fact that the set of observations has been obtained from different normal distributions, with different means and / or variances, and the mixture of all these data in a single data set does not have to follow a single normal distribution.

In these cases, the finite mixture technique of Gaussian distributions can be applied. It consists of decomposing the data into multiple normal distributions, estimating the mean and the corresponding variance, as well as the probability in the mixture. In this way, a population model can be generated that allows a better understanding of the nature of the analysed variable. In this work we present an application of this procedure for data from an ALS campaign flight in which errors were measured on different types of terrain.

Keywords: Normal distribution, Finite mixture models, EM estimation



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

CIENCIAS SOCIALES



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Análisis Multivariante de las políticas de cierre y contención efectivas frente al COVID-19: países de Europa

Joel Antonio Martínez-Regalado ⁽¹⁾, Cinthia Leonora Murillo-Avalos ⁽²⁾, Purificación Vicente-Galindo ⁽³⁾, J.L. Vicente-Villardón ⁽⁴⁾

(1) Departamento de Estadística, Universidad de Salamanca, e-mail: joel_martinez@usal.es

(2) Departamento de Estadística, Universidad de Salamanca, e-mail: cinthia_muav@usal.es

(3) Departamento de Estadística, Universidad de Salamanca, e-mail: purivg@usal.es

(4) Departamento de Estadística, Universidad de Salamanca, e-mail: villardon@usal.es

RESUMEN

El SARS-CoV-2 (COVID-19) prevalece desde el 2019, como la enfermedad de mayor cuidado en consecuencia de su fácil propagación y diagnósticos de fatalidad. En consecuencia, los gobiernos han respondido con la implementación de medidas de distanciamiento social con diferentes resultados. Por tanto, es necesario evaluar la eficacia de la aplicación de las políticas de cierre y contención sobre la tasa de infección y mortalidad. Utilizamos la técnica estadística multivariante Biplot Logístico Externo para analizar las medidas implementadas en los 19 países europeos con mayor número de casos. Los resultados obtenidos sugieren 3 clústeres según el nivel de rigurosidad implementado. Concretamente, el clúster con menor número de casos confirmados y defunciones combino la aplicación de cinco medidas, resultados ser el clúster con mayor nivel de rigor. Estos hallazgos deben ser considerados por los gobiernos, las organizaciones y la sociedad como una alternativa para prevenir potencialmente nuevas oleadas de COVID-19 o futuras pandemias.

Palabras clave: biplot, técnica estadística multivariante, políticas de contención y cierre, coronavirus, COVID-19, pandemias, SARS-CoV-2



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

SARS-CoV-2 (COVID-19) has prevailed since 2019, as the disease with the greatest care due to its easy spread and fatal diagnoses. Consequently, governments have responded with the implementation of social distancing measures with different results. Therefore, it is necessary to evaluate the effectiveness of the application of closure and containment policies on the infection and mortality rate. We used the statistical multivariate technique External Logistic Biplot to analyse the measures implemented in the 19 European countries with the highest number of cases. The results obtained suggest 3 clusters according to the level of rigor implemented. Specifically, the cluster with the lowest number of confirmed cases and deaths combined the application of five measures, resulting in the cluster with the highest level of rigor. These findings should be considered by governments, organizations and society as an alternative to potentially prevent new waves of COVID-19 or future pandemics.

Keywords: biplot, statistical multivariate technique, closure and containment policies, coronavirus, COVID-19, pandemic, SARS-CoV-2



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Responsabilidad social ambiental según el Global Reporting Initiative (GRI): un análisis multivariante de las empresas más grandes a nivel mundial

Cinthia L. Murillo-Avalos⁽¹⁾, Mitzi Cubilla-Montilla⁽²⁾, José Luis Vicente-Villardón⁽³⁾,
Purificación Vicente-Galindo⁽⁴⁾

(1) Departamento de Estadística, Universidad de Salamanca, e-mail: cinthia_muav@usal.es

(2) Departamento de Estadística, Universidad de Panamá, e-mail: mitzi.cubilla@up.ac.pa

(3) Departamento de Estadística, Universidad de Salamanca, e-mail: villardon@usal.es

(4) Departamento de Estadística, Universidad de Salamanca, e-mail: purivg@usal.es

RESUMEN

El interés por lograr un sistema social que cumpla con el desarrollo sostenible ha incrementado exponencialmente con los años, en primera causa por el notable deterioro de los ecosistemas, ocasionado por el mal uso de recursos, señalando a los ciudadanos corporativos como los principales actores. En consecuencia, las corporaciones han tenido que adecuar sus actividades empresariales en función de impactar positivamente en el medio ambiente. En este sentido, este trabajo tiene por objetivo conocer el compromiso que tienen las empresas más grandes a nivel mundial. Para lograrlo, hemos analizado el contenido de la dimensión ambiental de las memorias de sostenibilidad que generan las empresas ante el Global Reporting Initiative (GRI), aplicando el Biplot Logístico Externo (ELB), técnica estadística multivariante utilizada para datos binarios. Los resultados revelan que las mayores deficiencias en la publicación de información ambiental corresponden a las categorías de biodiversidad y mecanismos de denuncia ambientales. Asimismo, se identificó a las categorías de energía y emisiones como las más reportadas. Estudios similares son posibles de realizar para empresas que difunden sus informes de sostenibilidad bajo las directrices del GRI.

Palabras clave: biplot, técnica estadística multivariante, responsabilidad social ambiental, desarrollo sostenible, índice ambiental



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

The interest in achieving a social system that complies with sustainable development has increased exponentially over the years, primarily due to the notable deterioration of ecosystems, caused by the misuse of resources, pointing corporate citizens as the main actors. Consequently, corporations have had to adapt their business activities in order to positively impact the environment. In this sense, this work aims to understand the commitment of the largest companies worldwide. To achieve this, we analyse the content of the environmental dimension of sustainability reports generated by companies based on the Global Reporting Initiative (GRI), applying the External Logistics Biplot (ELB), a statistical multivariate technique used for binary data. The results reveal that the greatest deficiencies in the publication of environmental information correspond to the categories of biodiversity and environmental grievance mechanisms. Likewise, the energy and emissions categories were identified as the most reported. Similar studies are possible to carry out for companies that disseminate their sustainability reports under the GRI guidelines.

Keywords: biplot, statistical multivariate technique, environmental social responsibility, sustainability development, environmental index



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Técnicas para reducir el sesgo de selección en una encuesta online sobre adicciones

Ramón Ferri-García⁽¹⁾, María del Mar Rueda⁽²⁾, Francisca López-Torrecillas⁽³⁾

(1) Departamento de Estadística e Investigación Operativa, Universidad de Granada, e-mail: rferri@ugr.es

(2) Departamento de Estadística e Investigación Operativa, Universidad de Granada, e-mail: mrueda@ugr.es

(3) Departamento de Personalidad, Evaluación y Tratamiento Psicológico, Universidad de Granada, e-mail: fcalopez@ugr.es

RESUMEN

El desarrollo de nuevos métodos para la recogida de datos, como las encuestas online, ha sido particularmente ventajoso para la ciencia empírica en el sentido de reducir costes, así como aumentar la inmediatez y las posibilidades del cuestionario. Sin embargo, estos métodos están fuertemente afectados por el sesgo de selección, lo que conlleva estimaciones poco fiables. Algunos métodos han sido propuestos para reducir este sesgo de selección y obtener estimaciones más fiables, como la reponderación por calibración y el Propensity Score Adjustment (PSA). La eficacia de estos métodos está muy relacionada con la asociación existente entre la información auxiliar que empleen y las variables objeto de estudio; por lo tanto, se requiere seleccionar las variables auxiliares cuidadosamente. En este trabajo se realiza una aplicación de los métodos de calibración, PSA y selección automatizada de variables para mitigar el sesgo de selección en una encuesta online no probabilística sobre adicciones realizada al estudiantado de la Universidad de Granada. Utilizando información del censo de estudiantes y una encuesta probabilística de referencia, se obtienen estimaciones poblacionales de cada puntuación de adicción.

Palabras clave: encuestas online, sesgo de selección, Propensity Score Adjustment, calibración



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

The development of new data collection methods, such as online surveys, has been particularly advantageous for empirical sciences in terms of reduced costs, as well as enhancing immediacy and questionnaire possibilities. However, these methods are strongly affected by selection bias, leading to unreliable estimates. Some methods have been proposed to reduce this selection bias and obtain more reliable estimates, such as calibration reweighting and Propensity Score Adjustment (PSA). The efficiency of these methods is largely tied to the association between the auxiliary information used and the target variables; therefore, it is required to select auxiliary variables carefully. This study presents an application of calibration, PSA and automated variable selection to mitigate selection bias in a nonprobability online survey about addictions targeted to students from the University of Granada. Using information from the students' census and a reference probability sample, population estimates are obtained for each addiction score.

Keywords: online surveys, selection bias, Propensity Score Adjustment, calibration



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Análisis multivariante de la labor de la psicología en los Centros de Reconocimiento de Conductores

Armando González Sánchez⁽¹⁾

(1) Departamento de estadística, Universidad de Salamanca, armando_gonzalez@usal.es

RESUMEN

El trabajo de los psicólogos en los Centros de Reconocimiento de Conductores (CRC) presenta varias peculiaridades que intentan recogerse en este estudio. Se propuso una encuesta con algunos puntos considerados importantes para caracterizar a la población. Se ha obtenido un sondeo que mide las necesidades de los psicólogos en CRC, así como la situación de los CRC. A través de análisis multivariante se ha constatado una gran variabilidad en todas las dimensiones recogidas, menos en carga laboral, la cual es elevada. Los resultados muestran que se trata de un sector desunido organizacionalmente y expuesto a un marco de actuación inestable.

Palabras clave: Centros de Reconocimiento de Conductores, psicotécnico, psicofísico

ABSTRACT

The work of the psychologists in the Driver Recognition Centers (known by its Spanish acronym, CRC) presents several characteristics collected by this study. A survey was proposed with points that were considered important to characterize the population. A survey which was obtained, measures the needs of psychologists in CRC as well as the situation of CRC. A great variability has been found in all the dimensions collected except in high workload. The results show that it is a disjoint sector exposed to an unstable framework

Keywords: Driver Recognition Centers, psychotechnical, psychophysical



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA CIENTÍFICA

COMO



Universidad de Jaén

Escuela de Doctorado

Caracterización multivariante del uso de las TIC en estudiantes universitarios de Ciencias Sociales

María-Concepción Vega-Hernández⁽¹⁾, Carmen Patino-Alonso⁽²⁾, María-Purificación
Galindo-Villardón⁽³⁾

(1) Departamento de Estadística, Universidad de Salamanca, e-mail: mvegahdz@usal.es

(2) Departamento de Estadística, Universidad de Salamanca, e-mail: carpatino@usal.es

(3) Departamento de Estadística, Universidad de Salamanca, e-mail: pgalindo@usal.es

RESUMEN

La incorporación de las Tecnologías de la Información y la Comunicación (TIC) a los procesos de aprendizaje resulta fundamental en la Educación Superior. Sin embargo, cada alumno aprende de forma diferente y por ello utiliza las herramientas que mejor le convienen en su desarrollo académico.

Se ha realizado un análisis estadístico multivariante mediante el cuestionario REATIC utilizando una muestra (n=742) de estudiantes del área de Ciencias Sociales de la Universidad de Salamanca, para conocer el uso y la opinión que tienen sobre las TIC en su aprendizaje, así como caracterizarlos según la consideración de las TIC que utilizan.

Los resultados presentaron el manejo de los estudiantes sobre todo programas básicos y de interrelación personal, buscadores en la red y algunos navegadores web. Valoran la ayuda e importancia de las TIC en su proceso de aprendizaje y las consideran positivamente como herramientas relevantes.

Se identificaron cuatro clusters bien diferenciados identificados mediante la aplicación del análisis factorial para obtener una caracterización multivariante de los estudiantes: Uno formado por los estudiantes con un nivel medio de consideración y uso de las TIC, otro por los alumnos con un nivel bajo de consideración y uso de las TIC, un tercero formado por estudiantes con un nivel alto de consideración y uso de las TIC y un último formado por alumnos con un nivel alto de consideración de las TIC y un nivel medio de uso de las mismas.

Palabras clave: TIC, cluster, estudiantes, universidad.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA COMO HERRAMIENTA CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

The incorporation of Information and Communication Technologies (ICT) to the learning process is fundamental in Higher Education. However, each student learns in a different way and therefore uses the tools that best suit their academic development.

It has been a multivariate statistical analysis using the questionnaire REATIC using a sample (n=742) of students from the Social Sciences area at the University of Salamanca, to know the use and opinion they have about ICT in their learning, as well as to characterize them according to the consideration of the ICT they use.

The results presented students management above all basic and personal interrelation programs, search engines in the network and some web browsers. They value the help and importance of ICT in their learning process and consider positively them as relevant tools.

Four well-differentiated clusters were identified by applying the factor analysis to obtain a multivariate characterization of the students: One was formed by students with a medium level of ICT consideration and use, another by graduates with a low level of ICT consideration and use, a third formed by students with a high level of ICT consideration and use and a last one consisting of students with a high level of ICT consideration and a medium level of ICT use.

Keywords: ICT, cluster, students, University.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

PÓSTER



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Caracterización de la desigualdad de género. Una perspectiva multivariante usando HJ-Biplot

Nerea González-García ⁽¹⁾, Ana B. Nieto-Librero ⁽²⁾, Purificación Galindo-Villardón ⁽³⁾

(1) Departamento de Estadística, Universidad de Salamanca; Instituto de Investigación Biomédica de Salamanca (IBSAL), Salamanca, e-mail: nerea_gonzalez_garcia@usal.es

(2) Departamento de Estadística, Universidad de Salamanca, e-mail: ananieto@usal.es

(3) Departamento de Estadística, Universidad de Salamanca; Instituto de Investigación Biomédica de Salamanca (IBSAL), Salamanca, e-mail: pgalindo@usal.es

RESUMEN

En el año 2015 durante la Asamblea General de la Organización de las Naciones Unidas (ONU), los países miembros de las Naciones Unidas aprobaron la Agenda 2030 para el Desarrollo Sostenible. Se trata de un proyecto de actuación general, constituida por un conjunto de 17 Objetivos de Desarrollo Sostenible (ODS), continuación de los ocho objetivos de desarrollo del milenio, que agrupan 169 metas de aplicación universal, con desafíos referidos al desarrollo sostenible desde su enfoque social, económico y ambiental (pobreza, salud, cambio climático, educación,...). En ellos encuentra cabida el reconocimiento de la igualdad de género, centrada en los derechos humanos, y del empoderamiento de la mujer. De entre las metas definidas, cabe destacar el ODS-5 que alude a “Lograr la igualdad entre los géneros y empoderar a todas las mujeres y las niñas”. En este sentido, en este trabajo se analizará desde una perspectiva multivariante el comportamiento del Índice de Desigualdad de Género, introducido por el Programa de las Naciones Unidas para el Desarrollo para medir las desigualdades de género en tres aspectos importantes del desarrollo humano (salud reproductiva, empoderamiento, y participación en el mercado de trabajo) en distintas regiones del mundo. Para ello, se hará uso del HJ-Biplot, técnica multivariante de reducción de la dimensión que facilita la representación gráfica de observaciones y variables en un mismo espacio de dimensión reducida, donde las relaciones entre ambas son visualmente interpretables. El análisis se implementó en el software libre R, haciendo uso de la librería *biplotbootGUI*.

Palabras clave: Índice de Desigualdad de Género, Objetivos de Desarrollo Sostenible, HJ-Biplot



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

In 2015, during the General Assembly of the United Nations (UN), the countries of the United Nations approved the 2030 Agenda for Sustainable Development. It is a project for general action, consisting of a set of 17 Sustainable Development Goals (SDGs), extension of the eight Millennium Development Goals, which include 169 universally applicable goals, with challenges related to sustainable development from a social, economic and environmental perspective (poverty, health, climate change, education, etc.). They include the recognition of gender equality, focused on human rights, and the empowerment of women. Among the defined goals, it is important to highlight SDG-5, which refers to "Achieve gender equality and empower all women and girls". In this regard, this study analyzes from a multivariate perspective the trends of the Gender Inequality Index, introduced by the United Nations Development Program to measure gender inequalities in three important aspects of human development (reproductive health, empowerment, and participation in the labor market) in different regions of the world. For this purpose, HJ-Biplot was employed. HJ-Biplot is a multivariate dimension reduction technique that facilitates the graphical representation of matrix observations and variables in the same low-dimensional space, where the relationships between them are visually interpretable. The analysis was implemented in the free software R, using the *biplotbootGUI* library.

Keywords: Gender Inequality Index, Sustainable Development Goals, HJ-Biplot



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Estudio del Índice Desarrollo Humano a través del Clustering Disjoint HJ-Biplot

Ana B. Nieto-Librero⁽¹⁾, Nerea González-García⁽²⁾, Purificación Galindo-Villardón⁽³⁾

(1) Departamento de Estadística, Universidad de Salamanca; Instituto de Investigación Biomédica de Salamanca (IBSAL), Salamanca, e-mail: ananieto@usal.es

(2) Departamento de Estadística, Universidad de Salamanca, e-mail: nerea_gonzalez_garcia@usal.es

(3) Departamento de Estadística, Universidad de Salamanca; Instituto de Investigación Biomédica de Salamanca (IBSAL), Salamanca, e-mail: pgalindo@usal.es

RESUMEN

Cada vez es más frecuente en Minería de Datos encontrarnos con la necesidad de combinar técnicas dedicadas, por un lado, a la clasificación de objetos y, por otro lado, a la búsqueda de patrones mediante representaciones en baja dimensión, pero con una mejora de la interpretación. Recientemente, se ha propuesto un método denominado Clustering Disjoint HJ-Biplot en el que se da una solución al doble problema de clasificación y mejora de la interpretación de los resultados en un espacio de dimensión reducida. En este trabajo se mostrará la utilidad del CDBiplot mediante en un conjunto de índices que miden diferentes aspectos relacionados con el Índice de Desarrollo Humano (IDH). El IDH es un indicador del desarrollo humano de un país elaborado por el Programa de las Naciones Unidas para el Desarrollo (PNUD). Sintetiza los alcances de un país respecto de la salud, la educación y la riqueza. Por ello, partiendo de un conjunto de indicadores que miden las dimensiones mencionadas anteriormente en diferentes países, el objetivo de este trabajo es encontrar patrones de similitud entre ellos respecto a estas dimensiones y comparar los mismos con la clasificación de los países según su grado de Desarrollo Humano presentada por el PNUD. Para analizar esta base de datos y estudiar la estabilidad de los resultados utilizaremos el CDBiplot (Nieto-Librero et al, 2017) a través de la librería biplotbootGUI del entorno R.

Palabras clave: Índice de Desarrollo Humano, Clustering Disjoint HJ-Biplot, componentes principales sparse.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

Nowadays, it is increasingly common in Data Mining to combine object classification techniques and dimensionality reduction methods to search patterns in a low-dimensional space, but with an improvement in results' interpretation. Recently, a method called Clustering Disjoint HJ-Biplot has been proposed to solve the double problem of classification and improvement of the interpretation of the results in a space of reduced dimension. In this work the usefulness of CDBiplot is shown by means of a set of indexes that measure different aspects related to the Human Development Index (HDI). The HDI is an indicator of a country's human development developed by the United Nations Development Programme (UNDP). It summarizes a country's achievements in terms of health, education and standard of living. Therefore, the objective of this work is to find patterns of similarity between a set of indicators that measure the dimensions mentioned above in different countries and compare them with the classification of the countries according to Human Development Index ranking presented by the UNDP. To analyze this database and study the stability of the results we will use the CDBiplot, implemented in biplotbootGUI library of the R environment.

Keywords: Human Development Index, Clustering-Disjoint HJ-Biplot, sparse principal components.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA CIENTÍFICA

COMO



Universidad de Jaén

Escuela de Doctorado

Aplicación de técnicas estadísticas multivariantes y de reducción de dimensionalidad en variables de carga externa y contextuales en un equipo de fútbol semiprofesional.

Enrique Benítez Andrés⁽¹⁾, Mercedes Sanchez Barba⁽²⁾, Mario Sánchez García⁽³⁾, Javier Sánchez Sánchez⁽⁴⁾

(1) Departamento de Estadística, Facultad de Medicina, Universidad de Salamanca, e-mail: ebeneitez@usal.es

(2) Departamento de Estadística, Facultad de Medicina, Universidad de Salamanca, e-mail: mersanbar@usal.es

(3) Facultad de Educación, Universidad Pontificia de Salamanca, Grupo de Investigación "Planificación y evaluación del entrenamiento y rendimiento deportivo", e-mail: msanchezga@upsa.es

(4) Facultad de Educación, Universidad Pontificia de Salamanca, Grupo de Investigación "Planificación y evaluación del entrenamiento y rendimiento deportivo", e-mail: jsanchezsa@upsa.es

RESUMEN

Interés estudio: Nuestros hallazgos, basados en técnicas multivariantes y de representación gráfica a través de reducción de dimensionalidad, nos permiten aportar herramientas optimizadoras respecto a la preparación física teniendo en cuenta variables contextuales intrínsecas del fútbol. **Objetivos:** A partir del registro de la carga externa semanal estimar el proceso de entrenamiento para poder optimizar el rendimiento de los jugadores. **Material y Métodos:** Mediante un sistema de geoposicionamiento (GPS), se registraron las variables de carga externa como la distancia total (Drel), distancia a sprint (DSP), distancia a alta (DAV), media (DMV) y baja velocidad (DBV) y el número de aceleraciones (ACC), todas ellas expresadas en función del tiempo de participación, en un equipo de fútbol semiprofesional (3ª división) durante 7 microciclos consecutivos compuestos por cinco sesiones de entrenamiento categorizadas según el día del partido o matchday (MD). Los datos fueron analizados según la condición del equipo (local vs visitante) y según el rol del jugador (titular vs suplente). Para conseguir los objetivos propuestos se ha utilizado como técnica multivariante y de representación gráfica en un espacio de baja dimensión el HJ-Biplot (Villardón, M. P. G., 1986). **Resultados:** En la sesión postpartido (MD+1), la carga externa fue superior para todas las variables, tanto en suplentes como en titulares. Por otra parte, para las sesiones centrales de la semana (MD-4, MD-3 y MD-2) la carga externa fue superior cuando el equipo jugaba como local que cuando lo hacía como visitante lo que ha de tenerse en cuenta para la optimizar el rendimiento físico de los jugadores.

Palabras clave: HJ-Biplot, rendimiento, fútbol, monitorización GPS

Bibliografía del resumen:

Villardón, M. P. G. (1986). Una alternativa de representación simultánea: HJ-Biplot. *Qüestiió: quaderns d'estadística i investigació operativa*, 13-23.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Using a Bayesian Probabilistic Model and Machine Learning Tools for Smart Literature Review on Aquaculture Research

Javier De la Hoz-M⁽¹⁾, M^a José Fernández-Gómez⁽²⁾, Susana Mendes⁽³⁾

(1) Facultad de Ingeniería, Universidad del Magdalena, e-mail: jdelahoz@unimagdalena.edu.co

(2) Departamento de Estadística, Universidad de Salamanca, e-mail: mjfg@usal.es

(3) MARE - Marine and Environmental Sciences Center, ESTM, Polytechnic Institute of Leiria, e-mail: susana.mendes@ipleiria.pt

RESUMEN

Las revisiones exploratorias manuales de literatura deberían ser cosa del pasado, ya que la tecnología y el desarrollo de métodos de aprendizaje automático han madurado significativamente. Este estudio tuvo como objetivo descubrir los patrones de distribución y las tendencias de la literatura sobre acuicultura disponible para mejorar la comprensión, la naturaleza y la estructura de estas publicaciones. Este estudio realizó una revisión de la literatura de 38,319 resúmenes publicados en 14 revistas de acuicultura de primer nivel entre 1972 y 2019. Se aplicó la asignación de Dirichlet latente para realizar minería de texto en el conjunto de datos y se encontraron 40 tópicos clave. Se utilizaron herramientas de aprendizaje automático para la posterior distribución y composición de palabras. Como resultado, encontramos que el modelado de temas puede segregar una colección de artículos sobre diferentes temas y podría usarse como una herramienta para comprender la literatura, no solo recapturando hechos bien conocidos sino también descubriendo otros temas relevantes. En general, los temas confirman áreas clave de la investigación en acuicultura que han sido identificadas por estudios cualitativos; sin embargo, en nuestro caso, también proporcionan una evaluación y un análisis cuantitativo en la literatura científica más reciente.

Palabras clave: asignación latente Dirichlet; machine learning; modelado de tópicos; publicaciones en acuicultura; tendencia de tópicos



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

Exploratory manual reviews of the literature should be a thing of the past, as the technology and development of machine learning methods have evolved significantly. This study aimed to discover the distribution patterns and trends of the available aquaculture literature to improve the knowledge, nature and structure of these publications. This study performed a literature review of 38,319 abstracts published in 14 top-tier aquaculture journals between 1972 and 2019. A Latent Dirichlet allocation was applied to perform text mining on the dataset; we found 40 topics. Machine learning tools were used for the subsequent distribution and composition of words. As a result, we found that topic modeling can segregate a collection of articles on different topics and could be used as a tool to understand literature, not only recapturing well-known facts but also discovering other relevant topics. In general, the topics confirm key areas of aquaculture research that have been identified by qualitative studies. However, in our case, they also provide a quantitative evaluation and analysis in the most recent scientific literature.

Keywords: latent Dirichlet allocation; machine learning; topic modeling; aquaculture publications; topics trend



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Análisis multivariante del uso de las TIC y las estrategias de aprendizaje en futuros ingenieros

María-Concepción Vega-Hernández⁽¹⁾, Carmen Patino-Alonso⁽²⁾, Jose-Luis Molina⁽³⁾,
María-Purificación Galindo-Villardón⁽⁴⁾

(1) Departamento de Estadística, Universidad de Salamanca, e-mail: mvegahdz@usal.es

(2) Departamento de Estadística, Universidad de Salamanca, e-mail: carpatino@usal.es

(3) Departamento de Ingeniería Hidráulica, Universidad de Salamanca, e-mail: jlmolina@usal.es

(4) Departamento de Estadística, Universidad de Salamanca, e-mail: pgalindo@usal.es

RESUMEN

Los estudiantes de cualquier ingeniería deben ser capaces de gestionar sus propios recursos para aprender, adaptarse y resolver problemas. Por esta razón, los futuros ingenieros deben saber utilizar estrategias de aprendizaje adecuadas y Tecnologías de la Información y la Comunicación (TIC).

Se ha realizado un estudio mediante los cuestionarios ACRA-Abreviada y REATIC en una muestra (n=348) de estudiantes del área de Ingeniería de la Universidad de Salamanca, para conocer el uso y la opinión que tienen sobre las TIC, así como sus estrategias de aprendizaje.

Los resultados presentaron un predominio de hábitos de estudio en la mayoría de estudiantes de ingeniería, seguido de un uso de estrategias cognitivas y de control del aprendizaje y, por último, de estrategias de apoyo al aprendizaje. En cuanto a las TIC, se observó el gran manejo de programas básicos y de buscadores en la red como Google o Yahoo. Valoran la importancia de las TIC en su proceso educativo, la ayuda que suponen cuando deben buscar información y las consideran fáciles de comprender y utilizar. Además, se llevó a cabo un Biplot Logístico que permitiera estudiar en profundidad la presencia/ausencia de las variables en los ingenieros.

Palabras clave: ingenieros, TIC, estrategias, biplot logístico, universidad.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
CIENTÍFICA
COMO



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

Students of any engineering must be able to manage their own resources to learn, adapt and solve problems. For this reason, future engineers must know how to use appropriate learning strategies and Information and Communication Technologies (ICT).

It has been a study using the questionnaire Abridged ACRA and REATIC to a sample (n=348) of students from the Engineering area at the University of Salamanca, to know the use and opinion they have about ICT, as well as their learning strategies.

The results presented a predominance of study habits in the majority of engineering students, followed by the use of cognitive and control learning strategies and, lastly, learning support strategies. Regarding ICT, the great use of basic programs and search engines on the Internet such as Google or Yahoo was observed. They value the importance of ICTs in their educational process, the help they provide when they have to search for information and consider them easy to understand and use. In addition, a Logistics Biplot was carried out that allowed to study in depth the presence/absence of variables in the engineers.

Keywords: engineers, ICT, strategies, logistics biplot, university.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Comparación de las distribuciones EBW y Nielsen

José María Rodríguez Reinoso⁽¹⁾, Julia Rodríguez Reinoso⁽²⁾, Valentina Cueva López⁽³⁾

(1) Universidad de Granada, e-mail: jose.maria.rodriguez.reinoso@gmail.co

(2) Foqum analytics. e-mail: juliarodriguezreinoso@gmail.com

(3) Departamento de Estadística e Investigación Operativa, Universidad de Jaén, e-mail: vcueva@ujaen.es

RESUMEN

Desde el inicio del estudio de los modelos de distribuciones de probabilidad, se han desarrollado diversos modelos de probabilidad, así como se han generalizado en familias.

Si nos fijamos en el último año, el desarrollo de nuevos modelos para datos discretos no ha sido diferente al cómputo histórico de la Estadística. Por dicho motivo, en este trabajo se va a realizar el estudio comparativo de la distribución de Waring Biparamétrica Extendida y de la distribución de Nielsen. A través de la comparación de sus propiedades, de la utilización de la distancia de Kullback-Leibler así como de diversos ejemplos de aplicaciones a casos reales, queremos determinar cuándo es mejor utilizar un modelo que el otro.

Palabras clave: EBW, distribuciones, Nielsen, comparación.

ABSTRACT

Since the beginning of the study of probability distribution models, various probability models have been developed, as well as they have been generalized in families.

If we look at the last year, the development of new models for discrete data has not been different from the historical computation of Statistics. For this reason, in this work the comparative study of the Extended Biparametric Waring distribution and the Nielsen distribution will be carried out. Through the comparison of its properties, the use of the Kullback-Leibler distance as well as various examples of applications to real cases, we want to determine when it is better to use one model than the other.

Keywords: EBW, distribution, Nielsen, comparison.



Universidad de Jaén
Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén
Escuela de Doctorado

ESTADÍSTICA TEÓRICA



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA CIENTÍFICA

COMO



Universidad de Jaén

Escuela de Doctorado

Building optimization algorithms for high-dimensional robust linear regression

María Jaenada⁽¹⁾, Elena Castilla⁽²⁾

(1) Universidad Complutense de Madrid, mjaenada@ucm.es

(2) Universidad Complutense de Madrid, elecasti@ucm.es

RESUMEN

En las últimas décadas se han desarrollado varios métodos de regularización para modelos de regresión lineal para datos de alta dimensión. Los métodos de regularización realizan simultáneamente la selección del modelo y la estimación de los coeficientes de regresión, combinando una función de pérdida basada en los residuos y una función de penalización en los coeficientes de regresión que induce la selección de variables. Se han propuesto diferentes penalizaciones, entre las que se encuentran las penalizaciones LASSO o LASSO adaptativo, una variante que mejora la propiedad de selección del modelo, así como penalizaciones no cóncavas como SCAD o MCP, que producen estimadores menos sesgados. Sin embargo, la influencia de los valores atípicos es particularmente inconveniente en el contexto de datos de alta dimensión, por lo que deben considerarse métodos de estimación robustos. En este trabajo, se propone la utilización de medidas de divergencias como funciones de pérdida, incluyendo la pseudodistancia de Rényi (RP) y la divergencia de mínima potencia (DPD), para producir estimadores robustos. Así, se desarrollan algoritmos de optimización adaptados para las posibles combinaciones de funciones de pérdidas robustas y funciones de penalización, señalando sus diferencias y similitudes. Además, se examina la precisión en la estimación y los tiempos de ejecución de cada método propuesto mediante un estudio de simulación.

Palabras clave: Datos de alta dimension, modelo de regresión lineal, métodos de regularización, robustez.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
CIENTÍFICA
COMO



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

Over the last decades several regularization methods have been developed for sparse high-dimensional regression models. Regularisation methods simultaneously perform model selection and parameter estimation by merging a loss function based on the residuals and a penalty function on the regression coefficients that induces variable selection. Different penalties have been proposed, including LASSO or adaptive LASSO penalties, a variant that improves the selection property of the model, as well as non-concave penalties such as SCAD or MCP, which produce less biased estimators. However, the influence of outliers is particularly awkward in the context of high-dimensional data and thus, robust methods should be considered. In this work, we propose the use of divergence measures as loss functions, including the Rényi pseudodistance (RP) and the density power divergence (DPD), to produce robust estimators. Optimisation algorithms are developed for all possible combinations of robust loss functions and penalty functions, pointing out their differences and similarities. Furthermore, the performance and running times of each proposed method are examined through a simulation study.

Keywords: High-dimensional data, sparse linear regression model, regularization methods, robustness.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Uso del estado del arte de machine learning para hacer inferencia sobre una encuesta del COVID-19

Luis Castro Martín⁽¹⁾, Ramón Ferri García⁽²⁾ y María del Mar Rueda García⁽³⁾

(1) Departamento de Estadística e Investigación Operativa, Universidad de Granada, e-mail: luiscastro193@ugr.es

(2) Departamento de Estadística e Investigación Operativa, Universidad de Granada, e-mail: rferri@ugr.es

(3) Departamento de Estadística e Investigación Operativa, Universidad de Granada, e-mail: mrueda@ugr.es

RESUMEN

La importancia de los métodos online para obtener datos ha aumentado la relevancia de las encuestas no-probabilísticas, que dependen de mecanismos de auto-selección y sufren falta de cobertura. Estos problemas producen estimaciones sesgadas. Algunas técnicas como Statistical Matching y Propensity Score Adjustment han sido propuestas para compensar este sesgo usando una muestra probabilística auxiliar. Sin embargo, suelen aplicarse con regresión logística o lineal como algoritmo de machine learning. En este estudio, proponemos usar un algoritmo del estado del arte como es XGBoost. Se han llevado a cabo estudios de simulación que demuestran cuánto puede XGBoost mejorar los resultados actuales. Finalmente, mostramos cómo afecta esta metodología a las estimaciones de una encuesta sobre el COVID-19.

Palabras clave: encuestas no-probabilísticas, machine learning, covid-19

ABSTRACT

The importance of online methods for data collection has increased the relevance of non-probability surveys, since they depend on self-selection procedures and suffer from coverage problems. These issues result in biased estimations. Some techniques like Statistical Matching and Propensity Score Adjustment have been proposed in order to compensate this bias using an auxiliary probability sample. However, they usually rely on logistic or linear regression as the chosen machine learning algorithm. In this study, we propose using a state-of-the-art algorithm like XGBoost instead. Simulation studies are conducted to demonstrate how much can XGBoost improve the current results. Finally, we show how this methodology affects the estimations of a survey about COVID-19.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA CIENTÍFICA

COMO



Universidad de Jaén

Escuela de Doctorado

Keywords: nonprobability surveys, machine learning, covid-19

Biplot logístico usando algoritmos de machine learning

Jose Giovany Babativa-Márquez⁽¹⁾, Jose Luis Vicente-Villardón⁽²⁾

(1) Facultad de Ciencias de la Salud y el Deporte, Fundación Universitaria del Área Andina, e-mail:

jbabativa5@areandina.edu.co:

(2) Departamento de Estadística, Universidad de Salamanca, e-mail: villardon@usal.es

RESUMEN

Tener matrices de datos binarias es cada vez más frecuente, en este contexto el análisis de componentes principales para datos binarios viene siendo una técnica común para reducir la dimensionalidad y describir las relaciones existentes, pero no proporciona una representación simultánea entre variables e individuos. En este trabajo proponemos utilizar algoritmos de optimización multivariante para estimar los parámetros de un biplot logístico binario y visualizar de forma simultánea las filas y columnas de una matriz de datos binaria.

Se proponen cuatro algoritmos basados en el gradiente conjugado y se comparan con el algoritmo MM a través de un estudio de Monte Carlo, para evaluar el desempeño se consideran escenarios donde la matriz de datos tiene diferentes grados de imbalance, así como diferentes tamaños en las filas y en las columnas. Los métodos son comparados usando un error de entrenamiento y, también mediante un procedimiento de validación cruzada se estima el hiperparámetro que representa el número de dimensiones del modelo y se calcula el error de generalización. Los resultados muestran que los algoritmos propuestos tienen desempeños eficientes y encuentran correctamente la estructura de baja dimensión sin dejarse afectar por el grado de imbalance, ni tampoco se afectan por el número de filas o de columnas de la matriz binaria. Ilustramos el procedimiento utilizando datos binarios reales sobre una muestra de fallecidos por COVID-19, donde se aplican los métodos para realizar un análisis de la matriz binaria que representa las comorbilidades que tenían los pacientes.

Palabras clave: datos binarios, métodos de optimización, modelos de bajo rango, biplot logístico binario.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

Multivariate binary data is increasingly common, in this context, principal component analysis for binary data is one of the most used techniques for reducing dimensionality and describing existing relationships, but it does not provide a simultaneous representation of rows and columns. We propose to use multivariate optimization algorithms to estimate the parameters of a binary logistic biplot that allows the rows and columns of a binary data matrix to be represented simultaneously. To fit the model, the performance of four algorithms based on the conjugate gradient is evaluated and compared with the Majorization-Minimization algorithm through a Monte Carlo study, which considers scenarios with different degrees of imbalance, as well as different dimensions of the binary data set. A cross-validation procedure is used to select the hyperparameter that represents the number of dimensions in the model. The results show that the procedure based on the cross-validation error is successful to specify the model with the studied algorithms. Furthermore, the algorithms have similar performances and correctly find the low dimensional structure regardless of the degree of imbalance or the number of rows or columns. We illustrate the procedure using real binary data on a sample of deaths from COVID-19, where the methods are applied to perform an analysis of the binary matrix that represents the comorbidities that the patients had.

Keywords: binary data, optimization methods, low rank model, logistic binary biplot.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
COMO HERRAMIENTA
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

CIENCIAS DE LA EDUCACIÓN



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA CIENTÍFICA

COMO



Universidad de Jaén

Escuela de Doctorado

Determinantes de la competencia matemática de los estudiantes andaluces e implicaciones para el profesorado

David Molina-Muñoz⁽¹⁾, Antonio Jesús López-Montoya⁽²⁾, José Antonio Sánchez-Peigrín⁽³⁾, Francisco Martínez-Ortiz⁽⁴⁾, Elena Molina-Portillo⁽⁵⁾

(1) Departamento de Didáctica de la Matemática, Universidad de Granada, e-mail: dmolinam@ugr.es

(2) Departamento de Estadística e Investigación Operativa, Universidad de Jaén, e-mail: amontoya@ujaen.es

(3) Departamento de Didáctica de la Matemática, Universidad de Granada, e-mail: jpelegrin@ugr.es

(4) Departamento de Didáctica de las Matemáticas, Universidad de Jaén, e-mail: fmortiz@ujaen.es

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

RESUMEN

Los resultados en matemáticas de los estudiantes andaluces (467 puntos) en la edición de 2018 del informe PISA (*Programme for International Student Assessment*) distan mucho del promedio nacional (481 puntos). Este trabajo pretende identificar los factores que influyen en la competencia matemática del alumnado en Andalucía, a partir del análisis de la muestra andaluza de PISA 2018. Se evalúa el impacto de variables relativas al contexto personal y familiar de los estudiantes y a los centros a los que asisten. Dada la estructura anidada de los datos de PISA, se utilizará una metodología basada en la regresión multinivel para el análisis de la muestra andaluza, compuesta por 1766 estudiantes. Los resultados indican que la repetición de curso, la titularidad del centro, el género y el estatus socioeconómico y cultural influyen significativamente en la competencia matemática de los estudiantes. Por último, se discuten las implicaciones para el profesorado de secundaria de estos resultados.

Palabras clave: Andalucía, competencia matemática, evaluación educativa, PISA, profesorado.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

La Ciencia de datos en el currículo preuniversitario

José Antonio Sánchez Pelegrín⁽¹⁾, Felipe Ruz⁽²⁾, David Molina Muñoz⁽³⁾, Antonio Jesús López Montoya⁽⁴⁾, José Miguel Contreras García⁽⁵⁾

(1) Departamento de Didáctica de la Matemática, Universidad de Granada, e-mail: jpelegrin@ugr.es

(2) Departamento de Didáctica de la Matemática, Universidad de Granada, e-mail: felipe.ruz.angel@gmail.com

(3) Departamento de Didáctica de la Matemática, Universidad de Granada, e-mail: dmolinam@ugr.es

(4) Departamento de Estadística e Investigación Operativa, Universidad de Jaén, e-mail: amontoya@ujaen.es

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

RESUMEN

El auge de la información basada en datos debería venir acompañado de un mayor esfuerzo por mejorar la formación de los ciudadanos, proporcionando las herramientas necesarias para desenvolverse con éxito en la sociedad actual. Por este motivo se pretende abordar el desarrollo curricular en la Educación Obligatoria, concretamente a nivel preuniversitario, de una asignatura sobre Ciencia de Datos. Dicha propuesta se sustenta usando el marco teórico de las Estadísticas Cívicas, promoviendo las herramientas necesarias para obtener y/o analizar conclusiones, propias y/o ajenas, a partir de la recopilación y análisis de grandes conjuntos de datos.

En base a lo anterior, se ha llevado a cabo un juicio de expertos en la materia en el que participaron 12 especialistas españoles. El instrumento para estructurar las opiniones aportadas es un cuestionario en el que se requería aportar información sobre 4 ejes: la integración de la Ciencia de Datos en el currículo preuniversitario, los contenidos adecuados para el nivel, las herramientas para trabajar dichos contenidos y la metodología más idónea. Los resultados del estudio muestran que la mayoría de los expertos consultados apuestan por poner el énfasis en el análisis de datos y la interpretación de resultados en estos niveles, así como por el uso de metodologías activas que permitan aunar la enseñanza de los conceptos asociados con la Ciencia de Datos con un uso crítico de las TIC.

Palabras clave: Ciencia de Datos, educación preuniversitaria, Estadística Cívica.



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE ESTADÍSTICA HERRAMIENTA COMO CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

Análisis de las *fake news* relacionadas con el Coronavirus desde la perspectiva de las estadísticas cívicas

Francisco Martínez-Ortiz⁽¹⁾, Felipe Ruz⁽²⁾, Antonio Jesús López-Montoya⁽³⁾, José Antonio Sánchez Pelegrín⁽⁴⁾, David Molina-Muñoz⁽⁵⁾

(1) Departamento de Didáctica de las Ciencias, Universidad de Jaén, e-mail: fmortiz@ujaen.es

(2) Departamento de Didáctica de la Matemática, Universidad de Granada, e-mail: Felipe.ruz.angel@gmail.com

(3) Departamento de Didáctica de la Matemática, Universidad de Jaén, e-mail: amontoya@ujaen.es

(4) Departamento de Didáctica de la Matemática, Universidad de Granada, e-mail: jpelegrin@ugr.es

(5) Departamento de Didáctica de la Matemática, Universidad de Granada, e-mail: dmolinam@ugr.es

RESUMEN

Garantizar la participación activa de los ciudadanos en debates sobre aspectos de relevancia social requiere de la comprensión, interpretación y evaluación crítica de noticias que incluyen datos estadísticos. Este hecho requiere dominar técnicas estadísticas e 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 son las noticias que incluyen datos referentes al Coronavirus. El marco teórico utilizado es el de las *estadísticas cívicas*. Su estructura conceptual se organiza en tres dimensiones: la primera sobre aspectos afectivos, una segunda relacionada con cuestiones cognitivas y la tercera de competencias de comunicación. Los resultados indican la urgencia de incorporar conocimientos estadísticos especializados durante la enseñanza obligatoria y la formación profesional.

Palabras clave: Estadísticas cívicas, COVID-19, *fake news*, alfabetización estadística



Universidad de Jaén

Departamento de Estadística e
Investigación Operativa

IV JORNADAS DE
ESTADÍSTICA
HERRAMIENTA
COMO
CIENTÍFICA



Universidad de Jaén

Escuela de Doctorado

ABSTRACT

Guaranteeing the active participation of the citizens in debates related to relevant social aspects requires comprehension, interpretation and critical sense of the news that includes statistical data. Therefore, individuals, motivated by the commitment to interacting with real data and to communicating numerical evidence, should be able to use statistical techniques and be able to generate mental habits. Thus, in this paper we show some implications for citizenship when dealing with news that includes data related to Coronavirus. The theoretical framework used is *Civic Statistics*. Its conceptual structure is organised in three dimensions: the first is Engagement and Action, the second is Knowledge and the third is Enabling Processes. The results show the urgency to include specialised statistical knowledge during the compulsory and vocational education.

Keywords: Civic statistics, COVID-19, fake news, Statistical literacy