Book of Abstracts

IV International Workshop on Proximity Data, Multivariate Analysis and Classification

April 25-26, 2019

Salamanca (Spain)

Editors:

José Fernando Vera, Universidad de Granada.
Eva Boj del Val, Universidad de Barcelona.
José Luis Vicente Villardón, Universidad de Salamanca.
Laura Vicente González, Universidad de Salamanca.

Edited by the Publications Service of the University of Barcelona.
ISBN: 978-84-09-16359-5

General Index

Organizing Committee

José Fernando Vera, UGR (jfvera@ugr.es)
Eva Boj del Val, UB (evaboj@ub.edu)
José Luis Vicente Villardón, USAL (villardon@usal.es)
María José Fernández Gómez, USAL (mjfg@usal.es)

Scientific Committee

Carles Mª Cuadras, UB (ccuadras@ub.edu)
José Fernando Vera, UGR (jfvera@ugr.es)
Eva Boj del Val, UB (evaboj@ub.edu)
José Luis Vicente Villardón, USAL (villardon@usal.es)
Mª Purificación Galindo, USAL (pgalindo@usal.es)
Albert Santorra, UPF (albert.satorra@upf.edu)
J. A. Martín, UDG (josepantoni.martin@udg.edu)
Elías Moreno, UGR (emoreno@ugr.es)
Luis Ángel García, UVA (lagarcia@eio.uva.es)
Christian Hennig, UCL (c.hennig@ucl.ac.uk)

Important Dates

April, 20th: Deadline for abstract submission
April, 21th: Notification of acceptance
April, 25th: Deadline for registration

University of Salamanca, November 2019.

Location

Facultad de Medicina de la Universidad de Salamanca
Alfonso X el Sabio, s/n, 37007 Salamanca

Sponsors

http://www.seio.es/

http://www.usal.es

http://soce.iec.cat/

Annual meeting of the SEIO Working Group in Multivariate Analysis and Classification (AMyC), Salamanca, April 25-26, 2019

The IV International Workshop on Proximity Data, Multivariate Analysis and Classification will take place during April 25-26, 2019 in Salamanca (Spain):

http://biplot.dep.usal.es/waymc/

It is organized by the Multivariate Analysis and Classification Spanish SEIO Group AMyC:

http://amyc.seio.es/

The Spanish Group of Multivariate Analysis and Classification is a Working Group of more than 50 researchers from all the Spanish universities. Every year, the Working Group organizes a meeting to promote the communication between its members and between them and other researchers, and to contribute to the development of the Multivariate Analysis and Classification field and related problems and applications. The first International Workshop on Proximity Data, Multivariate Analysis and Classification took place in Granada, October 2014 (http://www.ugr.es/~amyc/EVENTS/WAMYC/), the second in Barcelona, October 2016 (http://www.ub.edu/wamyc/) and the third in Valladolid (http://www.eio.uva.es/wamyc/).

The topics of interest comprise any related problem to Multivariate Analysis and Classification both from a theoretical or a computational point of view, and their applications. It also includes problems related to unsupervised or supervised statistical learning related to big data analysis.

Reduced Program

<table>
<tr><th>Time</th><th>Thursday, 25th</th><th>Friday, 26th</th></tr>
<tr><td>9:00 – 10:00</td><td><u>Registration</u></td><td>–</td></tr>
<tr><td>10:00 – 10:30</td><td><u>Opening:</u>
Mª Purificación Galindo Villardón
José Luis Vicente-Villardón
Eva Boj</td><td rowspan="2"><u>Plenary Invited Talk:</u>
Chair: Agustín Mayo-Iscar
k-quantiles clustering
Christian Martin Hennig</td></tr>
<tr><td>10:30 – 11:00</td><td rowspan="2"><u>Plenary Invited Talk:</u>
Chair: José Fernando Vera
Hierical disjoint principal component analysis
Mauricio Vichi</td></tr>
<tr><td>11:00 – 11:30</td><td><u>Oral Session 2</u>
Chair: Elena María Castilla González
Attraction-repulsion clustering with applications to fairness
Hristo Inouzhe
One-shot device testing under exponential distribution:a new robust approach
Elena María Castilla</td></tr>
<tr><td>11:30 – 12:00</td><td colspan="2">Coffee break</td></tr>
<tr><td>12:00 – 13:30</td><td><u>Invited Talks</u>
Chair: José Luis Vicente-Villardón
On the equivalence of the GEM and Variational procedures for a latent block Poisson model
José Fernando Vera
Could logratio methods be useful to analise the generalized propensity score?
Marc Comas-Cufí
Redundancy analysis for categorical data based on logistic responses
José Luis Vicente-Villardón</td><td><u>Poster Session</u>
Chair: Eva Boj</td></tr>
</table>

13:30 – 16:00	**Lunch**	
16:00 – 17:00	**<u>Plenary Invited Talk:</u>** *Chair: Marc Comas-Cufí* Inference methods for multivariate latent variable models ALBERT SANTORRA	**AMyC Group Meeting**
17:00 – 18:00	**<u>Oral Session 1:</u>** *Chair: Luis Ángel García-Escudero* Obtaining fairness using optimal transport theory PAULA GORDALIZA Directional multivariate quantiles: definition, guidelines, implications and advantages RAÚL ANDRÉS TORRES Cluster analysis with cellwise outliers LUIS ÁNGEL GARCÍA-ESCUDERO	

Program

Thursday, April 25th, 2019

9.30-10 Registration (Salon de Grados - Facultad de Medicina)

10-10.30 Opening (Salon de Grados - Facultad de Medicina)

Multivariate analysis and classification group of the statistics and operational research society

- Mª Purificación Galindo-Villardón, Vice-Chancellor of Postgraduate, University of Salamanca
- José Luis Vicente-Villardón, Chair of the Organizing Committee, University of Salamanca
- Eva Boj, Coordinator of the SEIO-AMyC Group, University of Barcelona

10.30-11.30 Plenary Invited Talk (Salon de Grados - Facultad de Medicina)

Chair: José Fernando Vera

Hierarchical disjoint principal component analysis

INVITED SPEAKER: Mauricio Vichi, Sapienza University of Rome

CO-AUTHORS: Cavicchia, C., Vichi, M., Zaccaria, G.

ABSTRACT:

Dimensionality reduction has been often considered in the last years due to the use of Big Data. Frequently the process of reduction has a hierarchically nested form which can be represented with a graphical configuration of a tree. Leaves correspond to manifest indicators (MIs), i.e., the portfolio or scoreboard of observed variables, while internal nodes denote components (that is, linear or nonlinear combinations of MIs) which synthesize common information in the data. The root of the tree is a general indicator. The hierarchy is a property which can be attributed to a manifold of different phenomena, from most general to most specific, in which a more general level includes more specific concepts. In this paper starting from the data matrix X of size (n × J), corresponding to n objects and J quantitative variables, we propose a statistical model for hierarchical parsimonious disjoint dimensionality reduction. This new methodology induces a hierarchical parsimonious system of indicators, each one represented by a component. The hierarchy is defined starting from the DPCA solution with a predefined number of latent variables. The components which take shape in the hierarchy could entail the identification of theoretical concepts, that represent the intermediate levels of the operationalisation phase for the construction of a composite indicator. The model is estimated by using the LS method, optimizing a constrained quadratic problem. Optimal properties, such as uniqueness and identifiability, are investigated. Albeit the HDPCA problem is NP-hard, a coordinate descent algorithm is proposed, and it turns out to be computationally efficient in real case studies. The goodness of fit of the

hierarchical parsimonious trees can be computed to assess the quality of the hierarchical partitions.

Mauricio Vichi, Sapienza University of Rome

http://www.dss.uniroma1.it/it/dipartimento/persone/vichi-maurizio

11.30-12 Coffee Break (Aula 5 - Sótano - Facultad de Medicina)

12-13.30 Invited Talks (Aula 5 - Sótano - Facultad de Medicina)

Chair: José Luis Vicente-Villardón

On the equivalence of the GEM and Variational procedures for a latent block Poisson model

INVITED SPEAKER: José Fernando Vera, University of Granada

ABSTRACT

In the maximum likelihood estimation of latent block models for contingency tables, usually a variational EM algorithm is employed. Although in a general block clustering framework this is a very useful procedure to address computational problems that may arise in a general formulation using GEM, in this work we show a parameterization for the traditional GEM algorithm that makes the two likelihood approaches give rise to an The equivalent estimation procedure in this context. This results is illustrated in a context of latent block distance association models (LBDA).

Could logratio methods be useful to analise the generalized propensity score?

INVITED SPEAKER: Marc Comas-Cufí, University of Girona

CO-AUTHORS: Comas-Cufí, M., Mateu-Figueras, G., Martín-Fernandez, J.A., Blanch, J., Ramos, R.

ABSTRACT:

In observational studies the intervention group of a particular individual is not defined at random. In fact, intervention groups are closely related to the characteristics of its individuals.

Moreover, it tends to happen that those characteristics are also related to the appearance of certain outcomes of interest. Therefore, if our analysis is not conditioned by those characteristics (commonly called confounders), measuring the effect that a specific intervention group has to the outcome can be misleading by this selection bias.

A well-known approach to reduce such bias consist on performing the analysis on similar individuals (matching). In propensity score analysis, individuals are considered individuals with similar probability of being assigned to each intervention group. In the particular case of only two intervention groups, the most common matching algorithms reduce to the bipartite matching case, obtaining fast and optimal algorithms for the matchings. When more than two interventions groups exist, no default algorithm exists to perform the matching between individuals.

In this work, using well-known concepts from compositional data analysis based on log-ratios, we propose a natural way to define closeness between individuals. Moreover, we show how the Aitchison geometry can help to guide the construction of the matching between individuals of more than two intervention groups. We will compare our results with other studies using a simulation scheme proposed by Lopez and Gutman (2017).

REFERENCES:

- Lopez,M.J. and Gutman,R. (2017). Estimation of causal effects with multiple treatments: A review and new ideas. Statistical Science 32(3), pp. 432-454.

Redundancy analysis for categorical data based on logistic responses

INVITED SPEAKER: José Luis Vicente-Villardón, University of Salamanca

ABSTRACT:

Redundancy analysis (RDA) is one of the many possible methods to extract and summarize the variation in a set of response variables that can be explained by a set of explanatory variables. The main idea is to use multivariate linear regression to explain the responses as a linear functions of the explanatory and then use Principal Component Analysis (PCA) or a Biplot to visualize the result. When response variables are categorical (binary, nominal or ordinal), classical linear techniques are not adequate. Some alternatives as Distance Based RDA have been proposed in the literature. In this paper we propose versions of RDA based on generalized linear models with logistic responses. The natural visualization methods for the visualization of the proposed techniques are the *Logistic Biplots* recently proposed. The methods are illustrated with an application to real data.

13.30-16 Lunch

1

6-17 Plenary Invited Talk (Salon de Grados - Facultad de Medicina)

Chair: Marc Comas-Cufí

Inference methods for multivariate latent variable models

INVITED SPEAKER: Albert Satorra, Pompeu Fabra University

ABSTRACT:

This talk overviews the concepts and different estimation methods for the practice of

multivariate latent variable modeling within the framework of structural equation models (SEMs). The relative performance of different estimation methods, including the Bayesian sem analysis, will be undertaken from the practitioner perspective of achieving proper estimates and sample stability in large and nearly non-identified models. We illustrate the conceptual issues and methods using the context of multivariate panel data and multitrait-multimethod (MTMM) models for measurement quality studies. We discuss the Bou and Satorra (2018)'s long vs. wide modeling approaches to panel data analysis. We also present the Saris and Satorra's (2018, 2019) recent work on estimation using pooled data (EUPD) for split-ballot MTMM data.
See http://84.89.132.1/~satorra/ for the work mentioned.

Albert Satorra

http://www.econ.upf.edu/~satorra/

17-18 Oral session 1 (Salon de Grados - Facultad de Medicina)

Chair: Luis Ángel García-Escudero

Obtaining fairness using optimal transport theory

SPEAKER: Paula Gordaliza, Université Toulouse III – Paul Sabatier and University of Valladolid

CO-AUTHORS: del Barrio, E., Gamboa, F., Gordaliza, P., Loubes, J.M.

ABSTRACT:
Statistical algorithms are usually helping in making decisions in many aspects of our lives. But, how do we know if these algorithms are biased against a subpopulation? Fairness is generally studied in a probabilistic framework where it is assumed that there exists a protected variable, whose use as an input of the algorithm may imply discrimination. There are different definitions of fairness in the literature. In this paper we focus on two of them which are called Disparate Impact (DI) and Balanced Error Rate (BER). Both are based on the outcome of the algorithm across the different groups

determined by the protected variable. The relationship between these two notions is also studied. The goals of this paper consist in detecting the lack of fairness of a binary classifier and then trying to fight against the potential discrimination attributable to it. This can be done by modifying either the classifiers or the data itself. Our work falls into the second category and modifies the input data using optimal transport theory.

Directional multivariate quantiles: definition, guidelines, implications and advantages

SPEAKER: Raúl Andrés Torres, University of Valladolid

ABSTRACT:

The notion of multivariate quantiles has a lack of uniqueness due to the absence of a total order in $\mathbb{R}^d$. However, it is an important notion to provide

important assessments such as risk boundaries or outliers detection. A variety of extensions can be found in the literature, some of them are related to directions which is an important factor to be highlighted in the multivariate setting.

Moreover, in practice, in-sample and out-sample frameworks arise based on the amount of data at disposal and the required $\alpha-$level of the quantiles. Thus, directional multivariate quantiles are presented to offer the inclusion of directions along with their properties and estimation methods in both in-sample and out-sample frameworks.

Cluster analysis with cellwise outliers

SPEAKERS: Luis Ángel García-Escudero, University of Valladolid

CO-AUTHORS: García-Escudero, L. A., Rivera-García, D., Mayo-Iscar, A., Ortega, J.

ABSTRACT:

A robust clustering procedure resorting to cellwise trimming and subspace approximations is presented. In problems of moderate and high dimensions, cellwise trimming is more appealing than casewise trimming because discarding entire observations causes a great loss of valuable information. The proposed methodology is particularized in the case of functional clustering in such a way that only "pieces" of the observed curves (where the curve is particularly outlying) are trimmed. Simulated and real data sets are considered to illustrate the proposed approach.

Friday, April 26th, 2019

10-11 Plenary Invited Talk (Salon de Grados - Facultad de Medicina)

Chair: Agustín Mayo-Iscar

k-quantiles clustering

INVITED SPEAKER: Christian Martin Hennig, University College of London

ABSTRACT:

A new cluster analysis method, K-quantiles clustering, is introduced. K-quantiles clustering can be computed by a simple greedy algorithm in the style of the classical Lloyd's algorithm for K-means. It can be applied to large and high-dimensional datasets.

It allows for within-cluster skewness and internal variable scaling based on within-cluster variation.

Different versions allow for different levels of parsimony and computational e-fficiency. Although K-quantiles clustering is conceived as nonparametric, it can be connected to a fixed partition model of generalized asymmetric Laplace-distributions. K-quantiles clustering is consistent for its canonical functional, and it is shown that K-quantiles clusters correspond to well separated mixture components in a nonparametric mixture. A simulation study has been carried out that shows in what situations K-quantiles clustering is superior to existing clustering methods, and it is applied to some real data if time allows.

Christian Martin Hennig

http://www.homepages.ucl.ac.uk/~ucakche/

11-11.30 Oral session 2 (Salon de Grados - Facultad de Medicina)

Chair: Elena María Castilla González

Attraction-repulsion clustering with applications to fairness

SPEAKER: Hristo Inouzhe, University of Valladolid

CO-AUTHORS: del Barrio, E., Inouzhe, H., Loubes, J.M.

ABSTRACT:

In the framework of fair learning, we consider clustering methods that avoid or limit the influence of a set of protected attributes, S, (race, sex, etc) over the resulting clusters, with the goal of producing a fair clustering. For this, we introduce perturbations to the Euclidean distance that take into account S in a way that resembles attraction-repulsion in charged particles in Physics and results in dissimilarities with an easy interpretation. Cluster analysis based on these dissimilarities penalizes homogeneity of the clusters in the attributes S, and leads to an improvement in fairness. We illustrate the use of our procedures with both synthetic and real data.

One-shot device testing under exponential distribution: a new robust approach

SPEAKER: Elena María Castilla, Complutense University of Madrid

CO-AUTHORS: Castilla, E.M., Martín, N., Pardo, L., Balakrishnan, N.

ABSTRACT:
Classical inferential methods for one-shot device testing data from an accelerated life-test are based on maximum likelihood estimators of model parameters. However, the lack of robustness of maximum likelihood estimator is well-known. We develop new estimators and Wald-type tests for one-shot device testing by assuming an Exponential distribution as a lifetime model. Through a theoretical study and a Monte Carlo simulation, the suggested estimators and tests are presented as a robust alternative to the maximum likelihood estimators and the classical Wald tests based on them.

11.30-12 Coffee Break (Hall - Facultad de Medicina)

12-13.30 Poster Session (Hall - Facultad de Medicina)

Chair: Eva Boj

The provision R package for claims reserving

AUTHORS: Boj, E., Costa, T., Canadell, A. (University of Barcelona and Caixa Guissona)

ABSTRACT:
An R package has been built with RStudio for claims reserving using stochastic methods, to become an alternative tool for practitioners. The package has been developed in the framework of generalized linear models (GLM). Some deterministic models widely used by actuaries, as are the Chain-Ladder, the de Vylder's least squares and the Taylor's separation methods can be derived as particular cases. The package provides statistics and uncertainty measures of interest for reserves and allows to calculate the present value of future payments and the claims development results taking into account the Solvency II context.

Using Shiny as a tool for evaluation of health insurance pricing models

AUTHORS: Vargas, E., Boj, E., Costa, T. (University of Barcelona)

ABSTRACT:
The purpose in this work is the evaluation of real historical data of an Ecuadorian private health insurance company and the use of this information to provide a pricing model that considers many variables such as age, insured capital, deductible and sex. The research is specially focused on the study of the deductible since it works differently from traditional insurance. Generalized linear models are performed to predict both outpatient and inpatient care frequency and severity and the results are displayed in an interactive app using Shiny library. The link for the app is https://estebanvar90.shinyapps.io/TFM_app/. The app works from a server in Shinyapps.io and does not require to access to the computer that uploaded the app. All calculations are done in R using different libraries and personalized functions.

Claim frequency in home insurance and impact of macroeconomic variables

AUTHORS: Boj, E., Castañer, A., Claramnunt, M.M., Costa, T., Roch, O. (University of Barcelona)

ABSTRACT:
In this work the effect of the behavior of some economic factors on the loss ratio of home insurance is studied. It is fundamental for the insurer to foresee the impact that the evolution of the economic cycle can have on their business and appropriately modify the subscription cycle of the entity. The study is based on information from Spanish insurance companies provided by ICEA during the 2008-2015 period. A dynamic regression model is applied with the quarterly data of the claim frequency and includes indicators such as inflation, the unemployment rate or the evolution of the Gross Domestic Product as predictors.

Application of Multivariate Methods for Demographic Studies on Mortality in Ecuador

AUTHORS: Calderon-Cisneros, J., Bauz-Olvera, S., Pambabay-Calero, J., Robles-Amaya, J., Vicente-Villardón, J.L.

This work deals with the study of the distribution of the population in Ecuador with respect to its demographic structure, we reviewed the databases of general deaths from 1997 to 2017, all provinces of Ecuador, the number of variables present in the database is 67, based on our general objective we verified the data contained in the matrix of number of deaths (t) according to grouping of causes (List of main causes of death Becker), Period 1997 - 2017 taking the last 20 years, emphasizing from the perspective of multivariate analysis applying various models.

Water quality multivariate analysis: Application to Gatun Lake

AUTHORS: Carrasco, G., Patino-Alonso, M.C., Vicente-Galindo, M.P., Molina, J.L., Del C Castillo, M., Galindo-Villardón, M.P.

ABSTRACT:
The deterioration of water quality is a reason for concern about the disproportionate growth of the human population, the expansion of industrial and agricultural activity and the threats of climate change that may have important alterations in the hydrological cycle.

For this study, the measurement of ten quality variables was possible due to the availability of sophisticated instrumentation equipment for water quality measurement and analysis in the sampling sites, Gamboa and Paraíso, located in the Gatún lake, part of the Panama Canal Watershed.

Therefore, it is important through this work to highlight the use of multivariate statistical methods such as the HJ-Biplot that allows to inspect matrices of physical, chemical and biological data that have a mass of water at a given point and time. To identify the groupings, the different points of sampling and the affected variables, a cluster analysis has been conducted. The clusters were calculated through the Biplot-coordinates (K-means method, Euclidean distance).

The obtained results achieve the conformation of cluster of sampling points according to the months of the climatic season of the region. Reflecting the associations of physical-chemical and biological variables and pointing out the differences, where the collections of the samples were made. This analysis allowed us to identify the

physicochemical and biological variables that influenced the sets among the different sample points. Two cluster were found: cluster 1 formed by the sample points in Gamboa and Paraiso on December, February, March, April and May and cluster 2 formed by the sample points taken mainly in May at Paraiso, in July, August, September, October and November.

Application of the response theory to the item of the psychological well-being questionnaire of carol ryff

AUTHORS: Cortés-Rodríguez, M., Sánchez-Barba, M., Galindo-Villardón, M.P., Jarauta-Bragulat, E. (University of Salamanca)

ABSTRACT:
Carol Ryff's psychological well-being questionnaire is a measurement instrument that has been adapted to more than 20 countries and has numerous versions. The test has three versions created by the author, with 3 items, 7 items and 14 items for each dimension, and has a total of 6 dimensions.

In Spanish we have the adapted version of diaz et al (29 items), created from the Dutch version of Van Dierendonk (39 items). In this questionnaire the equality of the elements in the 6 dimensions of the test is not respected. Therefore, in this paper, we propose to start from the longest version of the Carol Ryff questionnaire, and create a new reduced version based on the evidence provided by the item response theory, trying to ensure that the dimensions are balanced.

For this, the response model has been used, which allows us to know which are the elements that provide information and the existence of response options that do not provide information.

Contributions to Biplot analysis: Disjoint solutions and sparse HJ-Biplot

AUTHORS: Cubilla-Montilla, M., Galindo-Villardón, P., Nieto-Librero, A.B., Vicente-Galindo, P., Torres-Cubilla, C.A.

ABSTRACT:ˇ
In the era of Big Data, vast amounts of data are being collected and curated in the form of arrays across the social, physical, engineering, biological, and ecological sciences. Big Data analysis relies on a variety of matrix decomposition methods which seek to exploit features exhibited by the high-dimensional data. With the massive growth of information (volume), the complexity of representation (variety) and the short time in which said information is stored (velocity), Big Data introduced a new problem to multivariate analysis, the interpretation problem.

In this work was considered three different algorithms to analyse and interpretation the results: HJ-Biplot, Disjoint Biplot and Sparse Biplot.

The **HJ-Biplot** (Galindo, 1986) is a technique which represents in a same graphics the variables and the observation. This is a considerable improvement to the GH Biplot and JK Biplot presented by Gabriel (1971).

The **Disjoint Biplot** (Nieto-Librero, Sierra, Vicente-Galindo, Ruíz-Barzola and Galindo-Villardón, 2017) is a technique which try to solve this problem. This forces the variables to provide information to only one component, facilitating the interpretation.

Here a new technique is presented: The **Sparse HJ Biplot**. This one is another alternative to solve the interpretation problem. *Sparse HJ Biplot* is proposed using *Elastic Net*

penalization. The size of the weights is based on a combination of l_1 norm and l_2 norm (Trendafilov, 2014; Zou, Hastie, 2003): $argmin_{\beta}\Sigma(Z_{i-}\boldsymbol{\beta}'\boldsymbol{x}i)2i+\lambda1\Sigma|\beta k|_{Kk=1}+\lambda2\Sigma\beta k2Kk=1$

REFERENCES:

- Gabriel, K.R., 1971. The biplot graphic display of matrices with application to principal component analysis. Biometrika 58, 453-467.
- Galindo, M.P. An alternative for simultaneous representation: HJ-Biplot Questíio, 10 (1986), pp. 12-23
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.
- Nieto-Librero, A. B., Sierra, C., Vicente-Galindo, M. P., Ruíz-Barzola, O., & Galindo-Villardón, M. P. (2017). Clustering Disjoint HJ-Biplot: A new tool for identifying pollution patterns in geochemical studies. Chemosphere, 176, 389-396.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B 58, 267-288.
- Trendafilov, N.T., 2014. From simple structure to sparse components: a review. Comput. Stat. 29, 431-454.
- Zou, H., Hastie, T., 2003. Regression Shrinkage and Selection via the Elastic Net, with Applications to Microarrays. Technical report. Department of Statistics, Stanford University.
- Zou, H., Hastie, T., Tibshirani, R., 2006. Sparse principal component análisis. J. Comp. Grapg. Stat. 15, 262-286

Alternatives to OLS in high-dimensional regression.

AUTHORS: García-Galindo, A., Torres-Cubilla, C., Sánchez-Barba, M. (University of Salamanca)

ABSTRACT:

High-dimensional regression is presented as a challenging task by the failure of Ordinary Least Squares In order to solve this problem, we perform a technical introduction over some of the most common statistical learning methods in the regression task and perform an empirical comparison between them on a high-dimensional gene expression dataset. We consider regularized linear models, tree-based models and dimensionality reduction methods. Boruta algorithm, which is based on Random Forest, was performed as a supervised feature selection method, bringing on us computational efficiency. To analyze the performace of the proposed models, different metrics were used to compare the prediction error committed by each model. The results show that most algorithms have a similar performance, although some significantly improve others. Finally, the models with the best performance are examples of solutions to solve the proposed problem and give an alternative to the Ordinary Least Squares.

Multivariate analysis reveals gene expression patterns altered in glioblastoma multiforme

AUTHORS: González-García, N., Nieto-Librero, A.B., Vital A., González-Tablas, M., Galindo-Villardón, M.P., Orfao, A., Tabernero, M. D. (University of Salamanca)

ABSTRACT:

Introduction. Gliomas are the most common brain tumors and astrocytic lineage, developed from glial cells, the 80% diagnosed brain neoplasias. According to the World

Health Organization (WHO) criterion, astrocytic gliomas are classified in non-diffuse and diffuse tumors ranked into four malignancy grades (from grade I to IV). Diffuse astrocytic gliomas group includes astrocytoma grade II (DAs, WHO II), anaplastic astrocytoma grade III (AAs, WHO III) and glioblastoma multiforme grade IV (GBMs, WHO IV). GBMs are the most agggresive and malignant form of astrocytomas, with high mortality rate due to their survival is lower than 5 years in most cases. These subtypes of diffuse astrocytic tumors are extremely heterogeneous and their prognosis and treatment are very diverse. Pathological diagnosis is the goal standard of practical clinical examination, yet the intratumoral and interobserver variabilities complicate the histological classification of the tumors. Thus, recognition of genetic alterations in diffuse astrocytic glioma subtypes could contribute to refine their current histological diagnosis.

Object. The goal of the present study was to focus in 3 diffuse astrocytic gliomas subtypes (DAs, AAs, GBMs) in order to identify the genetic alterations of GBMs, that differentiate them from lower grade astrocytomas.

Methods. We collect 176 astrocytoma samples (129 GBMs and 47 DAs and AAs) included in 7 different databases from GEO: one respective to our own cohort and 6 additional databases. Their RNA was processed using the Affymetrix HG-U133 Plus2 gene expression microarray chip. A total of 44723 probe sets, with information about 21336 genes, were kept in the analysis. Firstly, the 1000 probes with the greatest variability were selected using CUR decomposition and 26 differentially expressed genes were identified using ordinal penalized regression and foldchange values. Secondly, nonnegative matrix factorization was performed to represent discriminant gene expression patterns associated to diffuse glioma subtypes.

Results. These results led us to the discovery of a group of genes altered in GBMs. Malignant GBMs were principally distinguised from DAs&AAs by the overexpression of CHI3L1, IGFBP2, VEGFA, COL genes, NNMT, HOXD10, SHOX2, IGF2BP3 and MALSU1. We found ETNPPL, SH3GL2, PCDH7, SFRP2, DPP10, SNORC and FREM3 genes infraexpressed in GBMs. Angiogenesis and inflammatory cellular functions were specifically affected in GBMs.

Conclusions. This study determines a potentially group of discriminant genes for diffuse astrocytic gliomas classification, ilustrating the value of multivariate analysis in genomic characterization of GBMs. Future astrocytoma diagnosis should integrate novel molecular data, with outweigh histological features, combining genetic characteristics which correlate with patient malignancy and survival (CB16/12/00400 and ISCIII PI16/0476 grants).

Temporal stability analysis of the species-environment relationships in the region of continental Ecuador, using the STATICO muti-way method.

AUTHORS: González-Narváez, M., Fernández-Gómez, M.J., Mendes, S., Ruiz-Barzola, O., Galindo Villardón, M.P.

ABSTRACT:

The objective of this study was to identify the behavior pattern of the phytoplankton species when exposed to the environmental variations that occur in different seasonal conditions. For this purpose, the multivariate statistical method STATICO (Simier et al., 1999, Thioulouse, Simier and Chessel, 2004) was applied. Hence it was possible todetermine the co-structures between the pairs of tables environment-species and then to identify the stable part of that relationship. Therefore, the coast region of

continental Ecuador (Eastern Equatorial Pacific) was analyzed. The environmental paràmetres (temperature, dissolved oxygen, nitrate, nitrite, phosphate and silicate) and the phytoplankton data (23 species abundance) were collected at four sampling stations located ten miles offshore (Esmeraldas, Manta, La Libertad and Puerto Bolívar). Sampling was performed for seven standard different depths, monthly (February to December) and between the years of 2013 to 2015. It should be noted that the sampling period in analysis corresponded to different oceanographic conditions, namely the normal season (2013) and the extreme warm season (2015) in which an alteration in environmental and biological behavior occurred.

The results achieved from STATICO showed that the environment-species relationship was separated into two discriminatory groups. The first group highlights the relationship between the period between February and June (in particular, with a greater abundance overall). The second group showed the relationship that characterizes the period between July and December. It should be noted that this group is related with the classification of the months that make up the two climatic epochs of Ecuador, the rainy season comprised from December to May and the dry season (not very rainy) that goes from June to November (INOCAR, 2012). The period associated with the change of season, as well as the months (May, October, November and December) that precede and continue, presented the strongest environment-species co-structure.

REFERENCES:

- [1] INOCAR (2012) Capítulo I: Información general de la República del Ecuador Inocar 2012, Instituto Oceanografico de la Armada. Available at: http://www.inocar.mil.ec/web/index.php/derrotero-costas-ecuatorianas.
- [2] Thioulouse, J., Simier, M. and Chessel, D. (2004) 'Simultaneous analysis of a sequence of paired ecological tables', Ecology, 85(1), pp. 272–283.

Algorithms of selection and prediction multivariate for ordinal data

Authors: Martínez-Regalado, J.A., Montesinos-López, O.A. (University of Salamanca and University of Colima)

ABSTRACT:

Genomic-Selection offers the opportunity to accelerate the breeding cycle and increase grain production. Most genomic-enabled prediction models developed in this area work with continues variables and are normally distributed. Models that work with categorical scores for different traits and different environment interactions are lacking. For this reason, in this paper we proposed a Bayesian model for analyzing multiple traits and multiple environments for genome prediction modeled for ordinal data.

We use four different models; two of them with interactions, one with line X environment (L x E) and another one genomic X environment interaction (G x E) and genomic additive X genomic additive X environment interaction (G x G x E). We used one real data set to assess the predictive ability of genomic predictions for ordered categorical phenotypes. We applied R-Software package Bayesian Generalized Linear Regression (BGLR). We found that the models with interaction (model 1 and model 4) had better and worse predictive capacity respectively. We consider that it is necessary to test the models in other data sets to have results that can verify the ones already obtained.

Study of sustainability in Mexican companies based on the Global Reporting Initiative (GRI), applying the External Logistic Biplot

AUTHORS: Murillo-Avalos, C., Vicente-Galindo, P., Galindo-Villardón, M.P. (University of Salamanca and University of Colima)

ABSTRACT:
The planet has presented many changes for the pollution and human activities without conscience, that is why the term sustainable development is becoming more important every day. In the 70´s, some institutions began to report their activities that could affect the society. Actually the Global Reporting Initiative (GRI), is considered the leading organization worldwide in economic, environment and social sustainability. In 2015, it registered 7,579 companies, it is a non-profit organization that promotes the companies to generate their sustainability reports, to preserve the environmental and society.

The aim of this research is to determine the degree of the sustainability of Mexican companies and classify them according to their economic, social and environmental components through the External Logistic Biplot, multivariate method for the inspection of dichotomous data. The analysis was made to large, medium and small Mexican companies registered on the GRI website in 2011, 2012 and 2013.

It was found that the analyzed companies who provide more importance to sustainability index tend to evaluate labor practices and decent work, which is category of the social dimension.

After this study, it is concluded that similar studies can be carried out, that is, obtain the sustainability gradients of any company or country registered in the GRI website, applying the ELB.

Clustering disjoint HJ-Biplot: A new tool for identifying patterns in data mining

AUTHORS: Nieto-Librero, A.B., Sierra, C., González-García, N., Vicente-Galindo, M.P., Ruiz-Barzola, O., Galindo-Villardón, M.P.

ABSTRACT:
This work introduces a new mathematical algorithm termed Clustering Disjoint HJ-Biplot (CDBiplot), which searches for the underlying data structure in order to find the best classification of the object groups in a reduced space. To this end, disjoint factorial axes are generated, in which each variable only contributes to the formation of one factorial axis. A graphical representation of the individuals and variables is performed using the HJ-Biplot method. In order to facilitate the use of this new method within any practical context, a function in the language R has been developed. This work applies the CDBiplot to study an environmental geochemistry case involving environmental pollution in river Surface sediments. The study focuses on an area close to an important mining and metallurgical site, where sediments share a similar geological origin and chemical composition. The algorithm permitted a detailed study of the geochemical interactions and performed an excellent separation of the samples. Thus, the groups obtained were formed according to a similar geological origin, location, and nature of the anthropogenic inputs based only on chemical composition data. These results allowed clear identification of the sources of pollution and the delimitation of the polluted zones. All things considered, we conclude that the proposed algorithm is a powerful tool for studying environmental geochemistry data sets and suggest that the application of this methodology be extended to other research fields.

Hierarchical modeling and copulas for the study of diagnostic accuracy

AUTHORS: Pambabay-Calero, J., Bauz-Olvera, S., Nieto-Librero, A.B., Galindo-Villardón, M.P.

ABSTRACT:

Generally, in a diagnostic meta-analysis, summary measures such as sensitivity, specificity and odds ratio are used. However, they may not be suitable for integrating studies with different prevalence, different cut-off points, and heterogeneity into studies. The foregoing poses a problem to abstract measures. Therefore, before integrating a set of studies, it is necessary to study all these aspects in order to properly select the model.

Sensitivity and specificity measurement models that make more realistic assumptions about marginal distribution functions (e.g., those that deviate from the assumption of normality), although they may be more expensive in computational terms, are best measured by summaries of a meta-analysis of diagnostic tests. In this sense, models derived from copula offer a structure analytical flexible that is appropriate for the measurement of sensitivity and specificity.

The HSROC model does not operate on the original scale of sensitivity and specificity, but on the corresponding logit scale and by generally relying on the bivariate normal distribution for the random effects, it only allows one single correlation structure.

The use of copulas allows the study of dependencies with structures that are not necessarily linear which is possible in diagnostic situations whose results are obtained after dichotomization.

Multivariate profile of the e-participation based on x-statis method

AUTHORS: Rodríguez-Martínez, C.C., Galindo-Villardón, M.P., Nieto-Librero, A.B., García-Sánchez, I.M.

ABSTRACT:

The current status and evolution of the level of e-participation are analysed over the period 2008-2016, identifying the groups according to the degree of development of the countries present in the period and the factors that determine progress or regression what can happen. In contrast to previous studies, this research considers three types of determinants to explain the differences between the electronic participation development groups, with specific emphasis on political factors, institutional capacities and the organisational environment. The final sample consists of 101 countries.

The techniques of X-Statis and clustering are used to observe the relationships between e-participation and political characteristics, organisational environment and institutional capacities. X-Statis can be used to observe the evolution of individual indicators by year and country. Cluster analysis enables the grouping of individuals according to similar characteristics. Together, the methods allow us to observe the behaviour of individuals (countries) and variables (indicators) in a reduced space, and to obtain conglomerates.

Psychometric analysis of the assessment of the CaMir (Carles des Modeles Internes de Relations) based on The Classical Theory (CTT) vs Item-Response Theory (IRT)

AUTHORS: Sánchez-Barba, M., Cortés-Rodríguez, M., Tejedor-Valle, M., Vicente-Galindo, M.P. (University of Salamanca)

ABSTRACT:
In several branches of Science, investigators can run into the problem of the measurement when they try to carry out some researches; if there is not any “scale” able to assign an exact score, the work of the study turns out more complicated. These variables that are not immediately evident, are called latent constructs and just can be studied or deduced from indicators or variables measured meaning observable behaviors that allow to obtain an approximation as truthful as possible to the subject of study.

An appreciable number of measurement instruments appear from the interest of measure these conducts, and the questionnaire is one of these instruments; it is defined as a questions system ordered with coherence, with logical and psychological sense, expressed with a simple and clear language, which is replied in writing by the survey respondent generally without the pollster’s intervention (Córdoba, 2002). The questionnaire allows the recollection of primary sources information, that is, people who have the information of interest. It is the most used technique for collecting data in investigation because it is designed to quantify and generalize this information, and it is one of the less expensive technique. It is achieved a good questionnaire when it helps to obtain the necessary and sufficient information regarding the investigation’s purposes, and it maintains the respondent’s interest to get reliable and significant contents.

There are some theories that underlie the construction and analysis of tests and guide their building determining tests, what depends on the theoretical and statistical advances of each moment. These statistical theories are essential since they allow the psychometric properties estimation of tests and thus guaranteeing the decisions taken from them are appropriated.

There is two large standpoint or theories when it comes to build and analyze the tests, they are the Classical Test Theory (CTT) and the perspective of the Item Response Theory (IRT) (Fernández, 2010).

In the work we will carry aut a Psychometric analysis of the CaMir (Cartes des Modeles Internes de Relations), which allows to evaluate the internal models of attachment relations in the level of the representation elaboration more than the real experience (Marrone, 2001). CaMir consist of 72 items distributed into 13 scales, and it has been carried out CaMir adaptations to several countries population; recently it was elaborated a reduced version, CaMir-R, compound of 32 items shared out in seven dimensions, with format similar to Likert of five points with 1=Totally disagree and 5=Totally agree (Balluerka, Lacasa, Gorostiaga, Muela, & Pierrehumbert, 2011).

REFERENCES

- Balluerka, N., Lacasa, F., Gorostiaga, A., Muela, A., & Pierrehumbert, B. (2011). Versión reducida del cuestionario CaMir (CaMir-R) para la evaluación del apego. Psichothema,23 (3), 486-494.
- Córdoba, F. G. (2002). El cuestionario: recomendaciones metodológicas para el diseño de cuestionarios. Limusa.
- Fernández, J. M. (2010). Las teorías de los test: teoría clásica y teoría de respuesta a los ítems. Papeles del psicólogo, 31 (1), 57-66.
- Marrone, M. (2001). La teoría del apego: un enfoque actual. España: Psimática.

The structural consensus tree

AUTHORS: Soto-Becerra, R., Vicente-Villardon, J.L. (University of Salamanca)

ABSTRACT:
This paper presents a variant of the consensus tree: the structural consensus tree (SCT). This method combines the possibilities offered by traditional consensus tree methods with those of extending multidimensional metric scaling for three-way distance matrices, the DISTATIS (Abdi, 2009) which is a form of consensus k distance matrices. In this sense, SCT is presented as an improvement in the resolution of dendrograms obtained from consensus methods. In this work we explain the basic functioning of SCT from 28 data matrices containing in rows 72 countries and in columns 14 variables that characterize the countries by economic, environmental and socio-demographic descriptors, each of the matrices corresponds to one year from 1980 to 2007. The presentation that is made demonstrates the usefulness of this type of technique when one wishes to find a consensus classification based on data from three routes (individuals, variables and occasions). This work also includes stability measures to compare the configurations obtained from different cluster analysis strategies.

PLS Regression for detection of Black Sigatoka using Hyperspectral Images.

AUTHORS: Ugarte-Fajardo, J., Ochoa-Donoso, D., Criollo-Bonilla, R., Vicente-Villardón, J.L. (University of Salamanca)

ABSTRACT:
Black Sigatoka is the most devastating disease of banana plants in many countries around the world, which is caused by a fungal pathogen that affects the photosynthetic tissues of the banana leaves and degenerates chlorophyll production, resulting in changes in the structure of the leaves that lead to defoliation and early maturation of the fruit. By the time that symptoms are visible, the damages in banana plants are already irreversible.

Hyperspectral imaging (HSI) is a nondestructive analysis technology that offers speed, accuracy, and reliability, presenting a promising alternative for assessment of plants. Recent research realized in the Biotechnology Center (CIBE) and Robotic and Vision Center (CVR) of the ESPOL University, regarding the presence of black Sigatoka in Banana, have associated metabolic changes with pre-symptomatic signs of the disease. Specifically, the pathogen destroys the photosynthetic tissue, which induces an increase in fluorescence and heat emission in the leaf blade and modifies the transport pattern of the photoassimilates, affecting the production of chlorophyll due to necrotic or chlorotic lesions, that result in variations of reflectance in the regions VIS and NIR of the spectrum.

In this work, we use PLS, dimension reduction technique when the number of variables is greater than the number of observations complemented with a biplot representation (HS-Biplot) to classify healthy and infected leaves. We have also introduced a penalized logistic regression in the main algorithm to avoid separation problems that are likely to be present in our data. The application of both techniques allows, first, to reduce the number of variables to latent structures with the constraint that they explain as much as possible of the covariance between the predictor variables and the response variable, and then these latent variables are used to display individuals and their relationships with original variables by means of a biplot, which enables to perform a visual analysis of the structures of the data. These characteristics represent a significant contribution

to discriminate banana plants infected with the black Sigatoka fungus in its early stages and identify the wavelengths that are linked to the metabolic changes produced by the disease.

The PLS model accuracy using Leave-One-Out-Cross-Validation (LOOCV) was very high (0.98). The goodness-of-fit of the logistic regression is measured using some pseudo r-squared measures or the deviance and we got the best goodness-of-fit measures with a penalty equal to 0.1. The infected leaves were classified correctly, this includes to the asymptomatic inoculated leaves and infected leaves with severity 1 and 2.

The HS-Biplot graph shows the classification between healthy and infected leaves. The leaves with fewer infection areas are showing close to healthy leaves. The leaves with the highest level of infection, form a group located away from the healthy leaves and we can observe that the asymptomatic inoculated leaves are related to NIR wavelengths. This confirms that the disease, in its early stages, affects the structure of banana leaves and these changes, which are not yet visible, affect the reflectance captured by the spectrograph in the NIR wavelengths.

Analysis of Uruguayan adolescents work

AUTHORS: Urruticoechea, A., Vernazza, E. (Complutense University of Madrid)

ABSTRACT:

In the world, 152 million children and adolescents, between 5 and 17 years old, need to work to survive. The international regulation establishes that any work done by a In this research we analyze this last group, adolescents between 15 and 17 years old, from Uruguay. Specifically, the data comes from the Encuesta Nacional de Trabajo Infantil (ENTI) carried out in Uruguay between the months of September 2009 and May 2010. This survey revealed information about 935 adolescents, from 15 to 17 years old who perform some type of work, through a short form (questionnaire) with a structure of 6 sections associated with the following dimensions:

- Education (C)
- Domestic tasks (D)
- Economic activity (E)
- Occupational health and safety (F)
- Job search (G)
- Begging situations (H)

These dimensions will be considered as analysis variables (the dimension "job search" will not be considered in this work). In addition, there are variables: sociodemographic, age, sex and region of residence.

The main objective of this research is to characterize the work of Uruguayan adolescents between 15 and 17 years old. From this general objective, the following specific objectives arise: compare the work done by sex, analyze the level of work by origin (Capital vs. Interior) and by category (Child Labor vs. Non-Child Labor). For this, in this investigation, first a univariate descriptive analysis is carried out taking into account the sociodemographic variables. This preliminary analysis is complemented by a bivariate analysis of the dimensions. Finally, a multivariate analysis of the dimensions is carried out using the HJ Biplot methodology. Among the results obtained, stand out:

The majority of working adolescents (64.5%) are men. Of the total of adolescent workers, more than 65% are from the interior of the country. Of the total adolescent

workers, more than 17% do so under conditions of child labor, that is, they work more than 37 hours a week and/or perform risky jobs.

With respect to the dimensions under study, the dimension Education correlates positively with the dimension Household tasks and negatively with the dimension labor health and safety. The dimension economic activity correlates positively with the dimension occupational health and safety. It should be noted that the dimension of begging situations does not correlate linearly with any of the other dimensions.

Taking into consideration the dimensions and their behavior in relation to sex and place of residence, and their differentiation between CL and NCL, it is concluded that:

1. There are differences of means by sex in the dimensions Domestic tasks, economic activity, occupational health and safety.
2. There is only difference of means by place of residence in dimension Domestic tasks.
3. There is a difference in means, by CL and NCL in all dimensions except in situations of begging.

Finally, through the implementation of a HJ-Biplot can observed that:

1. The dimension of education correlates positively with domestic tasks (and negatively with economic activity and occupational health). In addition, the dimension of domestic activities correlates positively with the economic activity dimension and is independent of the occupational health and safety dimension, and the dimension of begging situations is independent of the rest of the dimensions.
2. The highest scores in the dimensions Economic activity and Occupational health and safety correspond to males, mostly from the interior and in the CL situation; the highest scores in the dimensions associated with Education and Homework correspond to women, mostly from the interior and in the situation of NCL.

As a final consideration, it is highlighted that these results go in the same direction as the studies previously carried out by UNICEF and by the Instituto Nacional de description.

Multivariate analysis of the learning styles and strategies' structures in university students

AUTHORS: Vega-Hernández, M.C., Patino-Alonso, M.C., Galindo-Villardón, M.P. (University of Salamanca)

ABSTRACT:

Nowadays, university students must acquire knowledge and skills to be able to successfully face any professional challenge that may arise. However, because of their experiences and skills each student learns in their own way. The most possible thing is for a student belonging to the branch of Science to learn and help himself from different techniques to an Art and Humanities' student, not only because of his way of being but because of his different way of learning in that area of knowledge. Some authors define the form or process in which a person accesses knowledge as a learning style, however, it also uses different strategies that help them in this task, called learning strategies. Therefore, to understand and improve student learning is essential to know this information together, so it is important to give a multivariate approach.

The objective of this research is to analyze the relationship between the structures of learning styles and strategies in university students in each areas of knowledge taking into account gender and academic qualification. For this, the COSTATIS method (Thioulouse, 2011) was applied, which consists of a co-inertia analysis (Doléc & Chessel, 1994) of the commitment of two k-table analyzes; that is, a consensus is established on the relationships between the variables of two data cubes, and only then builds a consensus for the areas, thus maximizing the co-inertia between the styles scores and learning strategies but looking for a description of the evolution of the relations and not of the stable part.

The study was carried out on a sample composed of 1979 students from the University of Salamanca (Spain), generating two series of matrices with the average total scores of the dimensions of Honey-Alonso Learning Styles Questionnaire (CHAEA) by Alonso, Gallego & Honey (1995) and Abridged ACRA Scale of Learning Strategies (ACRA-A) by De la Fuente & Justicia (2003), classifying the students by gender and academic grade from the previous course. So, the data were formed by I = 210 observations, J = 7 dimensional variables and K = 5 matrices that correspond to the areas of knowledge (Arts and Humanities, Sciences, Health Sciences, Social and Legal Sciences, and Engineering and Architecture).

The results showed 86.54% of inertial absorption explained on axis 1 and an acceptable correlation between structures (RV coefficient = 0.770). The co-structure graph showed that the values in the groups of women are similar in learning styles and learning strategies, nevertheless, in those of the men more discrepancies are observed. The co-inertia graph of matrices with the dimensions of learning styles and strategies presented associations between reflective style and study habits, and between pragmatic style and cognitive strategies and control of learning. The support strategies were related to the theoretical and pragmatic style, and an inverse relationship was observed between the active style and the study habits. The knowledge of this information allows teachers to help students improve their learning and optimize their time and resources.

A graphical representation associated to permanova for binary data

Authors: Vicente-González, L., Vicente-Villardón, J.L. (University of Salamanca)

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 arising, for example, from the treatments of a designed experiment, multivariate rather than univariate separate analysis should be used in order to control the Type I risk. The most popular method for multivariate comparisons is Multivariate Analysis of Variance (MANOVA) that can be considered as a Multivariate General Linear Model (MGLM). Normally, MANOVA is used together with a pictorial representation of the group centroids (Canonical Analysis) in order to help with the interpretation when the hypothesis of no group differences is rejected. The problem with MANOVA is that has very restrictive conditions for its correct application, namely, data has to be multivariate normal and the structure of variation and covariation must be the same across groups, moreover, the number of variables has to be much smaller the number of individuals. In many application fields as Genomics, in which the expression of a large number of genes is measured in only a few individuals, none of the previous conditions

holds and it is necessary to use non-parametric methods. We use PERMANOVA (MANOVA trough permutations) as an alternative to MANOVA when the application conditions do not hold. PERMANOVA was developed in Ecology to hold their highly asymmetric data but has not been extensively used in other fields as Genomics. In this poster we describe PERMANOVA as the main technique to study, with reference and the pictorial representations that can be associated to PERMANOVA, mainly Principal Coordinates Analysis on the group centroids, initially described in the Statistical literature but never applied to data. We will extend the technique initially described for numerical data to binary data. Finally, we will apply the main theoretical results to two sets of data, one from the NCI60 project that studies 60 cell lines from 10 different kinds of cancer and another one from the HapMap Project that studies single nucleotide polymorphisms in the human genome and its relation with health and other characteristics.

Multivariate characterization of students of the University of Salamanca

AUTHORS: Zárate-Santana, Z.J., Patino-Alonso, M.C., Sánchez-García, A.B., Celestino-Sánchez, M.A. (University of Salamanca)

ABSTRACT:

Introduction: In recent years numerous studies have contributed to boost the research on learning approaches and academic-stress coping by university students, evincing that the stress makes them change the way in which they learn. On the one hand, each student has a predetermined manner of approaching his/her learning. It can be superficial or deep, according to the degree of the involvement of students in their study determined by their creativeness, work capability and initiative to gather information beyond what professors provide (Hernández-Pina, García-Sanz, & Maquilón-Sánchez, 2005). On the other hand, students can cope with stress in three different fashions: positive re-evaluation (also known as cognitive reappraisal), searching for social support, and planning. The positive re-evaluation covers the strategy of focusing on the positive aspects and ignoring the negative ones of the problem. Students can also decide to seek social support (e.g., in lecturers, colleagues and/or family) in order to improve their emotional situation. Finally, planning consists of devising all the strategies to comprehend the problem and obtain an action plan to control the stress situation (Cabanach, Valle, Rodríguez, Piñeiro, & Freire, 2010).

Objective: Determine the relationship between the learning approaches and the academic-stress coping of the students of the University of Salamanca.

Instruments: Two questionnaires are employed. First, the questionnaire R-SPQ-2F, which is composed of 20 items with Likert-scale-like answers, is designed to measure the learning approach of students in their current teaching context. It is therefore an instrument to evaluate teaching rather than characterising students as "surface learners" or "deep learners" (Biggs, Kember, & Leung, 2001). Second, the academic-stress coping is quantified by the questionnaire A-CEA, which is a subscale of the questionnaire CEA (Spanish initials for "Questionnaire for Academic Stress") consisting of 23 items. It is formulated to evaluate the cognitive strategies and behaviors that allow the students to face the situations of academic stress, discerning between cognitive reappraisal, searching for social support and planning (Cabanach et al., 2010).

Methods: The Canonical Correspondence Analysis (CCA) through the programme CANOCO is used to relate the composition of the student community with the known variation in the university environment (Ter Braak, 1986).

Sample: 1,012 students of several knowledge areas of the University of Salamanca participated in the questionnaires.

Results: The dimensions of the questionnaires are confirmed in our sample and we find a correlation between learning approaches and academic-stress coping. Results revealed that, when students are under academic stress, social-support searching is preferred by surface learners while planning is the favorite strategy of deep learners.

13.30-16 Lunch

16-18 AMyC Group Meeting / Closing (Salón de Grados - Facultad de Medicina)

Invited Session of the AMyC-SEIO Working Group in the IFCS2019:

Wednesday, 28th August 2019

16.10 – 17.10 SP12: New developments in clustering and scaling data, organized by J. F. Vera & E. Boj del Val – Chair: Vicente-Villardón (Room I)

A cellwise trimming approach to Cluster Analysis, Luis Angel Garcia-Escudero, Diego Rivera-Garcia, Joaquín Ortega, Agustin Mayo-Iscar

Redundancy analysis for categorical data based on logistic regressions, Jose Luis Vicente-Villardon, Laura Vicente-Gonzalez

A log-ratio approach to cluster analysis of count data when the total is irrelevant, Marc Comas-Cufi, Josep Antoni Martin-Fernandez, Glòria Mateu-Figueras, Javier Palarea-Albaladejo

Index work

List of participants

Name	University	Correo electrónico
Agustín Mayo-Iscar	UVA	agustinm@eio.uva.es
Alar Urruticoechea	UCM	alarurru@ucm.es
Albert Satorra	UPF	albert.satorra@upf.edu
Alberto García-Galindo	USAL	alberto_26796@usal.es
Ana Belén Nieto Librero	USAL	ananieto@usal.es
C.A. Torres Cubilla	USAL	carlos_t22@usal.es
C.L. Murillo-Avalos	USAL	cinthia_muav@usal.es
Carles Mª Cuadras	UB	ccuadras@ub.edu
Carmelo C. Ávila Zarza	USAL	caaz@usal.es
Carmen Patino Alonso	USAL	carpatino@usal.es
Carmen Rodriguez	USAL and University of Panama	cc.rodriguezm@usal.es
Christian Martin Hennig	UCL	christian.hennig@unibo.it
Elena María Castilla	UCM	elecasti@ucm.es
Elena Vernazza	USAL	elenavernazza@usal.es
Eva Boj	UB	evaboj@ub.edu
Gilda Judith Taranto Vera	USAL	gilda.taranto@usal.es
Hristo Inouzhe	UVA	hristo.inouzhe@uva.es
Humberto Mauriciort Argotty Erajo	USAL	mauricio.argotti@usal.es
J. A. Martín-Fernández	UdG	josepantoni.martin@udg.edu
J.A. Martínez-Regalado	USAL	joel_martinez@usal.es
J.L. Vicente-Villardón	USAL	villardon@usal.es
Javier Martín Vallejo	USAL	jmv@usal.es
Johny J. Pambabay	ESPOL (ECUADOR)	jpambabay@usal.es
Jorge Ugarte-Fajardo	USAL	jugartef@usal.es
José Fernando Vera	UGR	jfvera@ugr.es
Juan Calderón Cisneros	UNEM (ECUADOR)	jtcalderon@usal.es
Julio Ernesto Salazar Pozo	USAL	julio_salazar@usal.es
L.A. García-Escudero	UVA	lagarcia@eio.uva.es
Laura Vicente-González	USAL	laura20vg@usal.es
M.C. Vega-Hernández	USAL	mvegahdz@usal.es
M.P. Galindo-Villardón	USAL	pgalindo@usal.es
Marc Comas-Cufí	UdG	marc.comas@udg.edu
María Cortés-Rodríguez	USAL	mariacortes@usal.es
Mª José Fernandez Gómez	USAL	mjfg@usal.es
Mariela González Narvaez	USAL	marielagn@usal.es

Mauricio Vichi	Sapienza University of Rome	maurizio.vichi@uniroma1.it
Mercedes Sánchez-Barba	USAL	mersanbar@usal
Miguel Marqués de Sousa	USAL	miguels1111@gmail.com
Mitzi Cubilla	USAL	mitzi@usal.es
Nerea González-García	USAL	nerea_gonzalez_garcia@usal.es
Paula Gordaliza	Université Toulouse III Paul Sabatier and UVA	paula.gordaliza@math-univ.toulouse.fr
Raúl Andrés Torres	UVA	raulandres.torres@uva.es
Rodrigo Soto-Becerra	USAL	rodrigosoto@usal.es
Sergio Bauz	ESPOL (ECUADOR)	sbauz@usal.es
Z.J. Zárate-Santana	USAL	zaira_zarate@usal.es
Elías Moreno	University of Granada	emoreno@urg.es

Photo Gallery


GRUPO AHTC

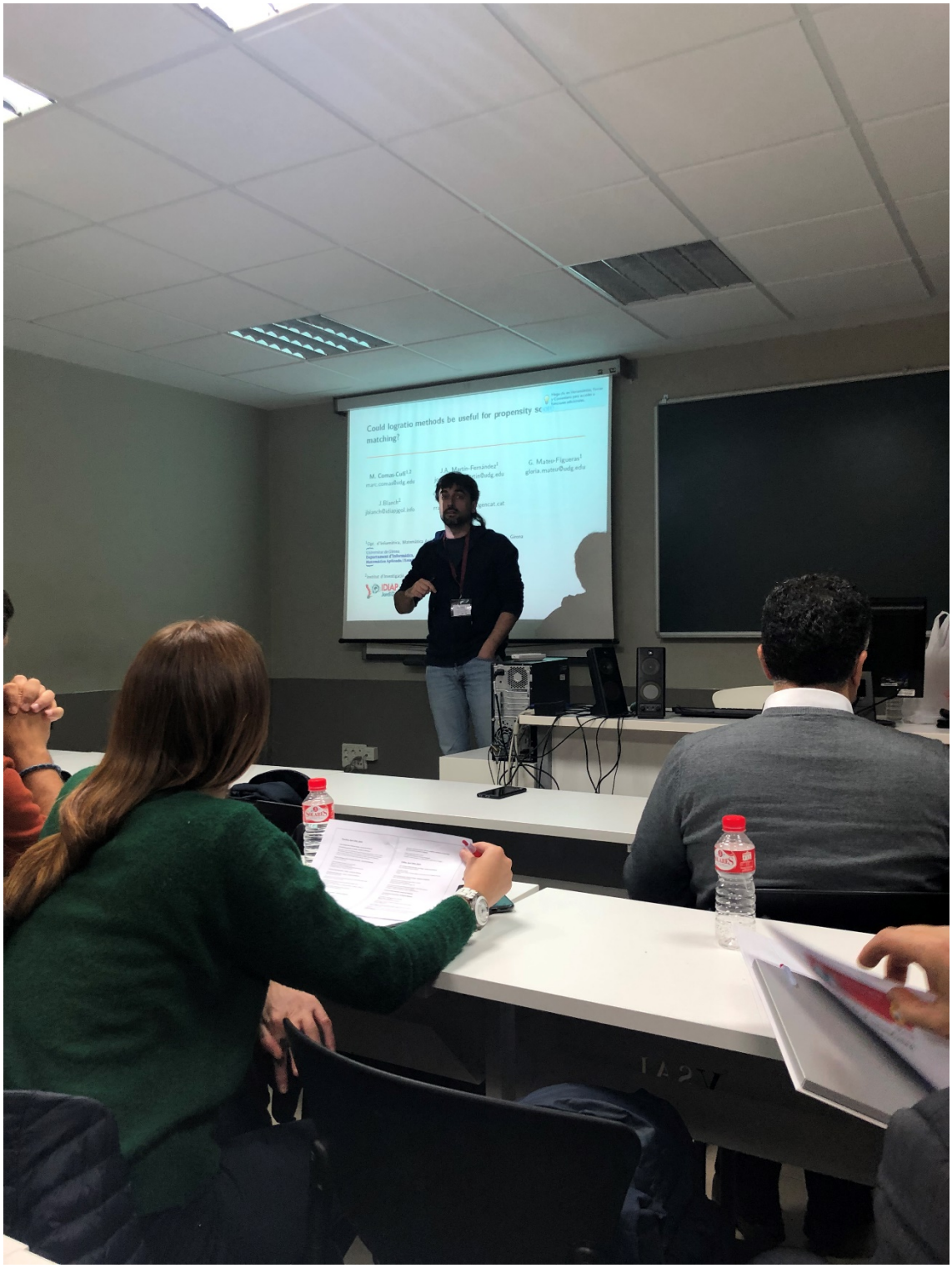

GRUPO AHTC


PRESENTATION
Next events


Measurement Model
Reflective Model
Formative Model
A
B


Organizations profiles vs intention to vote. IDPES 20


Organizations profiles vs intention to vote. (DPES 20


Could logratio methods be useful for propensity sc
matching?
G. Mateo-Figueras[1]
gloria.mateo@udg.edu


Could logratio methods be useful for propensity sc
matching?
M. Comas-Cufí[1,2]
marc.comas@udg.edu
J.A. Martín-Fernández[1]
josepantoni.martin@udg.edu
G. Mateu-Figueras[1]
gloria.mateu@udg.edu
J.Blanch[2]
jblanch@idiapjgol.info
R.Ramos[2]
rramos.girona.ics@gencat.cat
[1]Dpt. d'Informàtica, Matemàtica Aplicada i Estadística. Universitat de Girona. Girona
Universitat de Girona
Departament d'Informàtica,
Matemàtica Aplicada i Estadística
[2]Institut d'Investigació en Atenció Primària IDIAP Jordi Gol. Barcelona
IDIAP
Jordi Gol


Summary

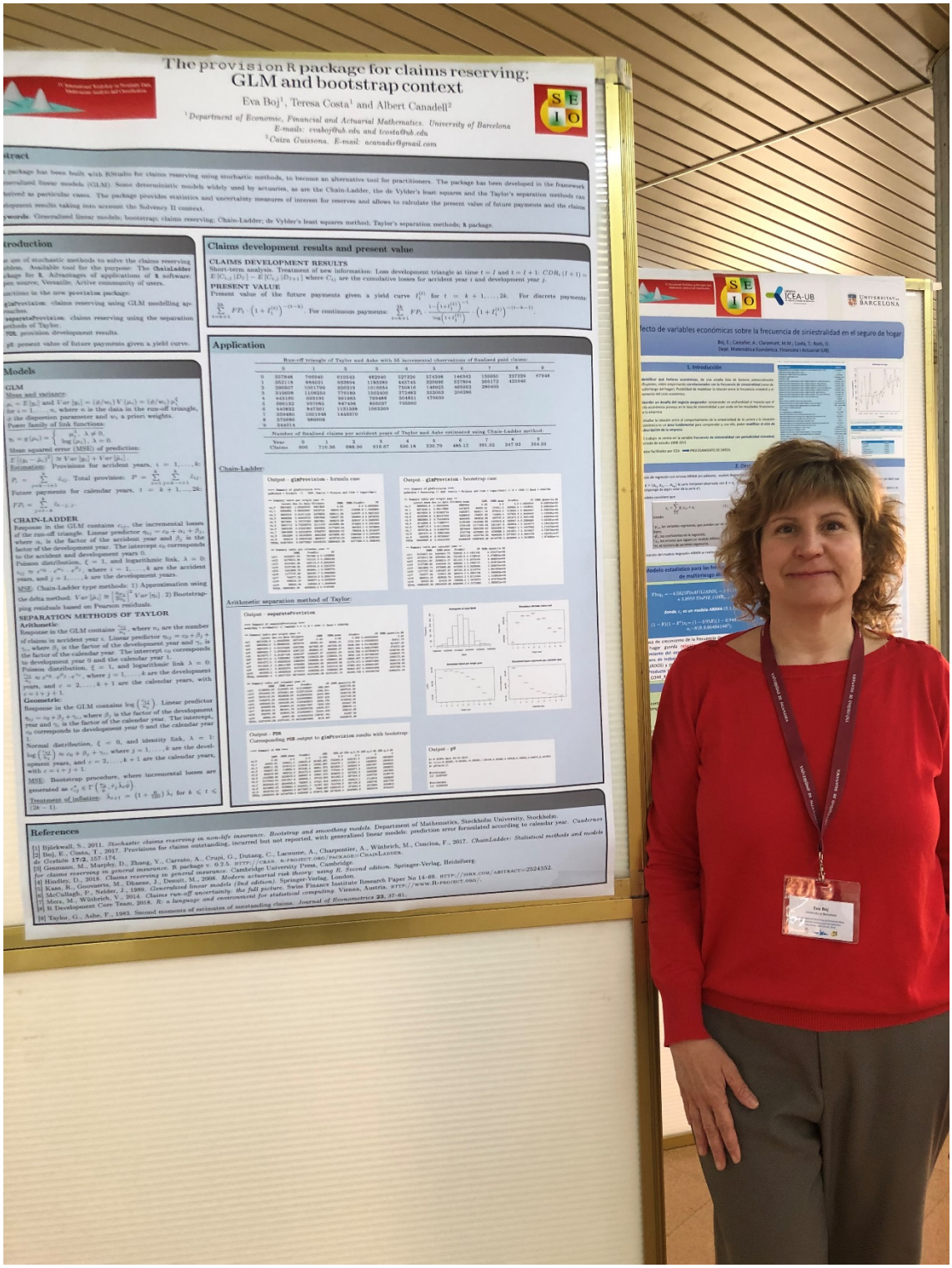

The Practice of Multivariate Latent Variable Models: Comparative of Methods
Albert Satorra
Universitat Pompeu Fabra
Barcelona, Spain
The Practice of Multivariate Latent Variable Models: Comparative of Methods
Albert Satorra
Universitat Pompeu Fabra
Barcelona, Spain

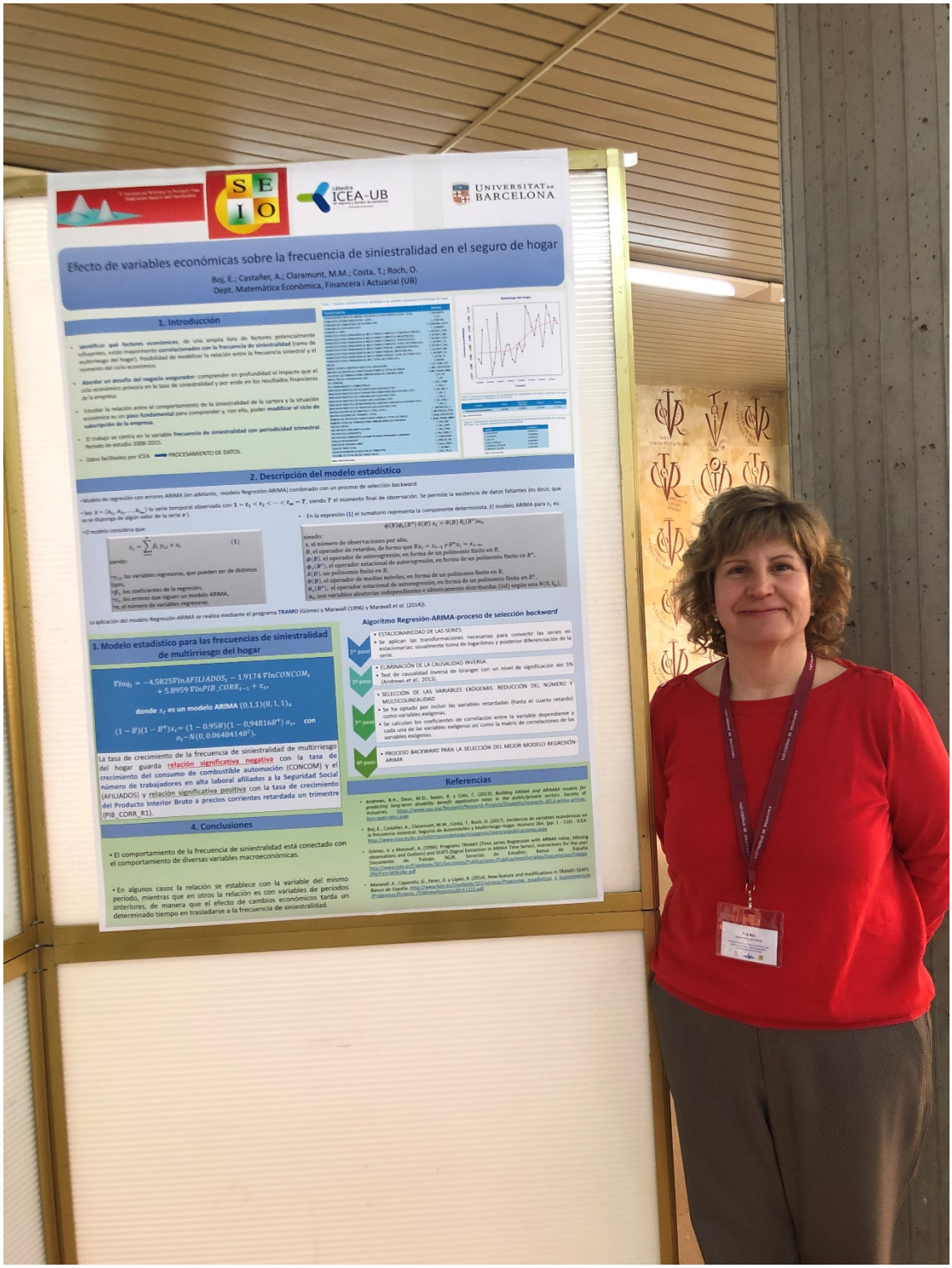

ifcs


ifcs

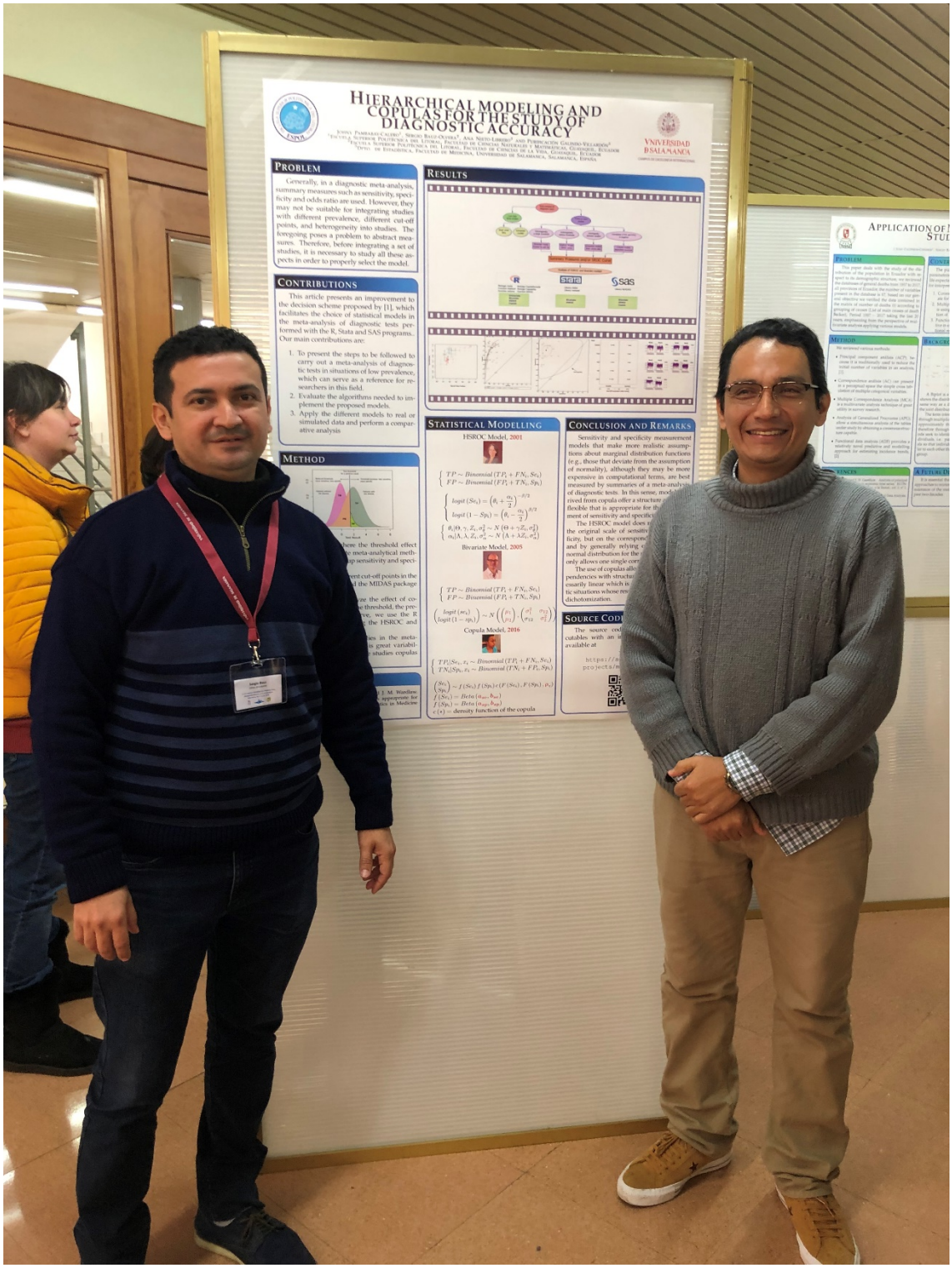

ifcs
CLUSTER BENCHMARK
DATA REPOSITORY


Using Shiny as a tool for evaluation of health insurance pricing models
Abstract
Outpatient and inpatient
Deductible
Claim frequency
Severity
Shiny app development
Final display
Coverage modifications:
Per loss Deductible
Coinsurance
Deductible for aggregate loss
Outpatient and inpatient analysis:
Claim frequency
Total loss:
Loss paid by the company
Deductible
Coinsurance
Generalized linear modeling

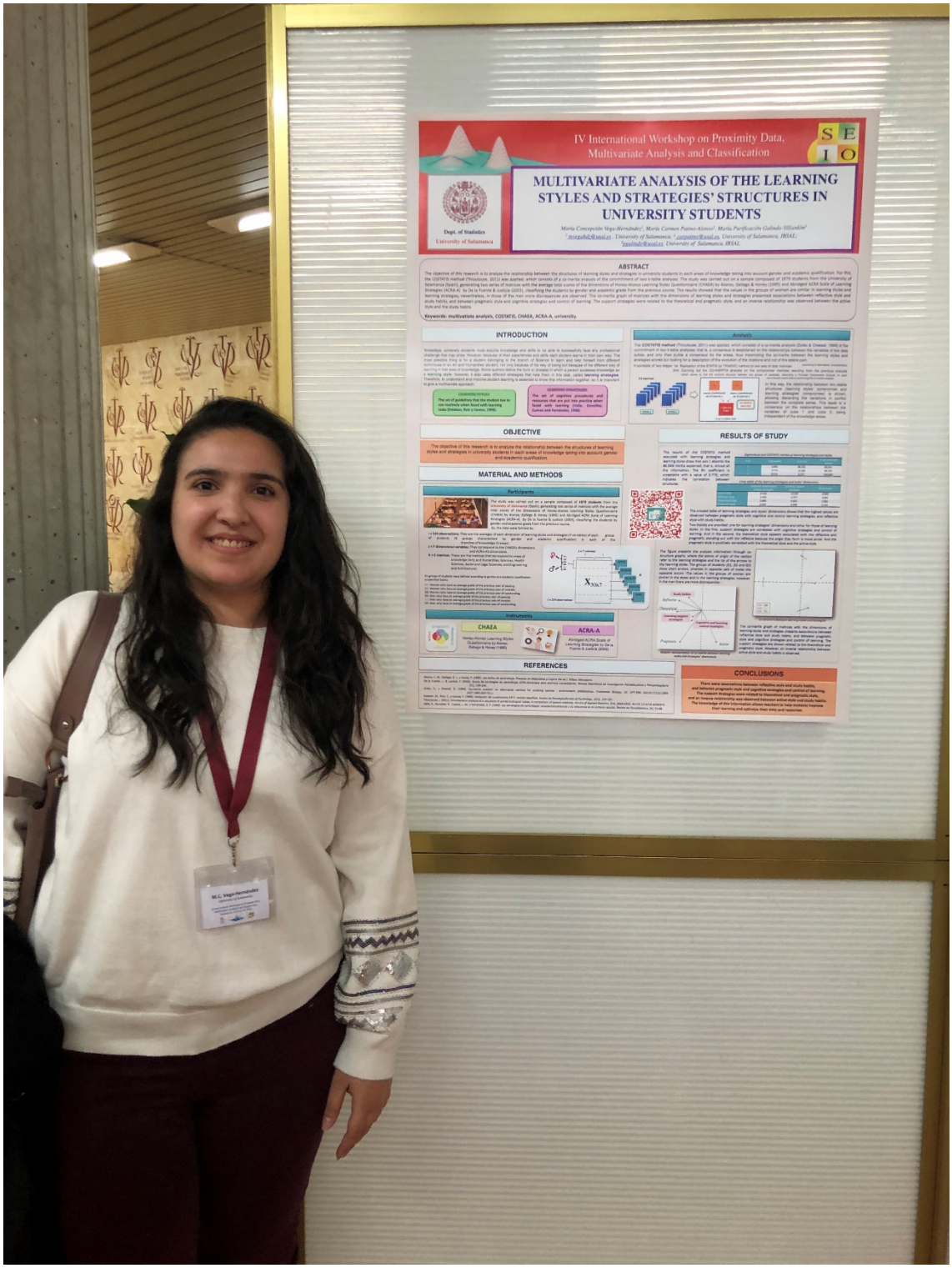

The provision R package for claims reserving: GLM and bootstrap context
Eva Boj[1], Teresa Costa[1] and Albert Canadell[2]
[1]Department of Economic, Financial and Actuarial Mathematics. University of Barcelona
E-mails: evaboj@ub.edu and tcosta@ub.edu
[2]Catra Guissona. E-mail: acanadis@gmail.com
Abstract
Introduction
Models
GLM
CHAIN-LADDER
SEPARATION METHODS OF TAYLOR
Claims development results and present value
CLAIMS DEVELOPMENT RESULTS
PRESENT VALUE
Application
Chain-Ladder:
Arithmetic separation method of Taylor:
References

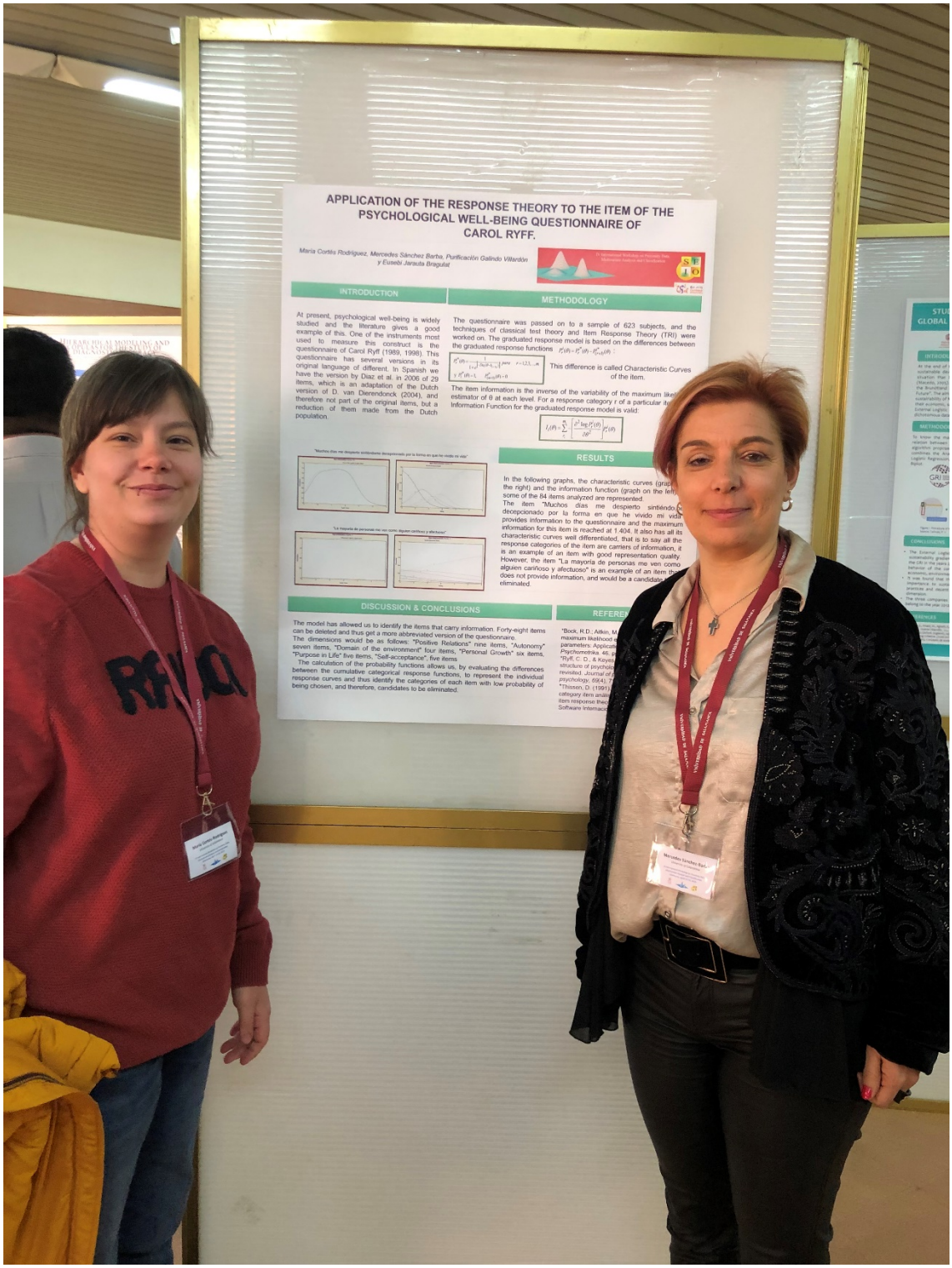

Efecto de variables económicas sobre la frecuencia de siniestralidad en el seguro de hogar
Boj, E.; Castañer, A.; Claramunt, M.M.; Costa, T.; Roch, O.
Dept. Matemàtica Econòmica, Financera i Actuarial (UB)
1. Introducción
2. Descripción del modelo estadístico
3. Modelo estadístico para las frecuencias de siniestralidad de multirriesgo del hogar
4. Conclusiones
Referencias

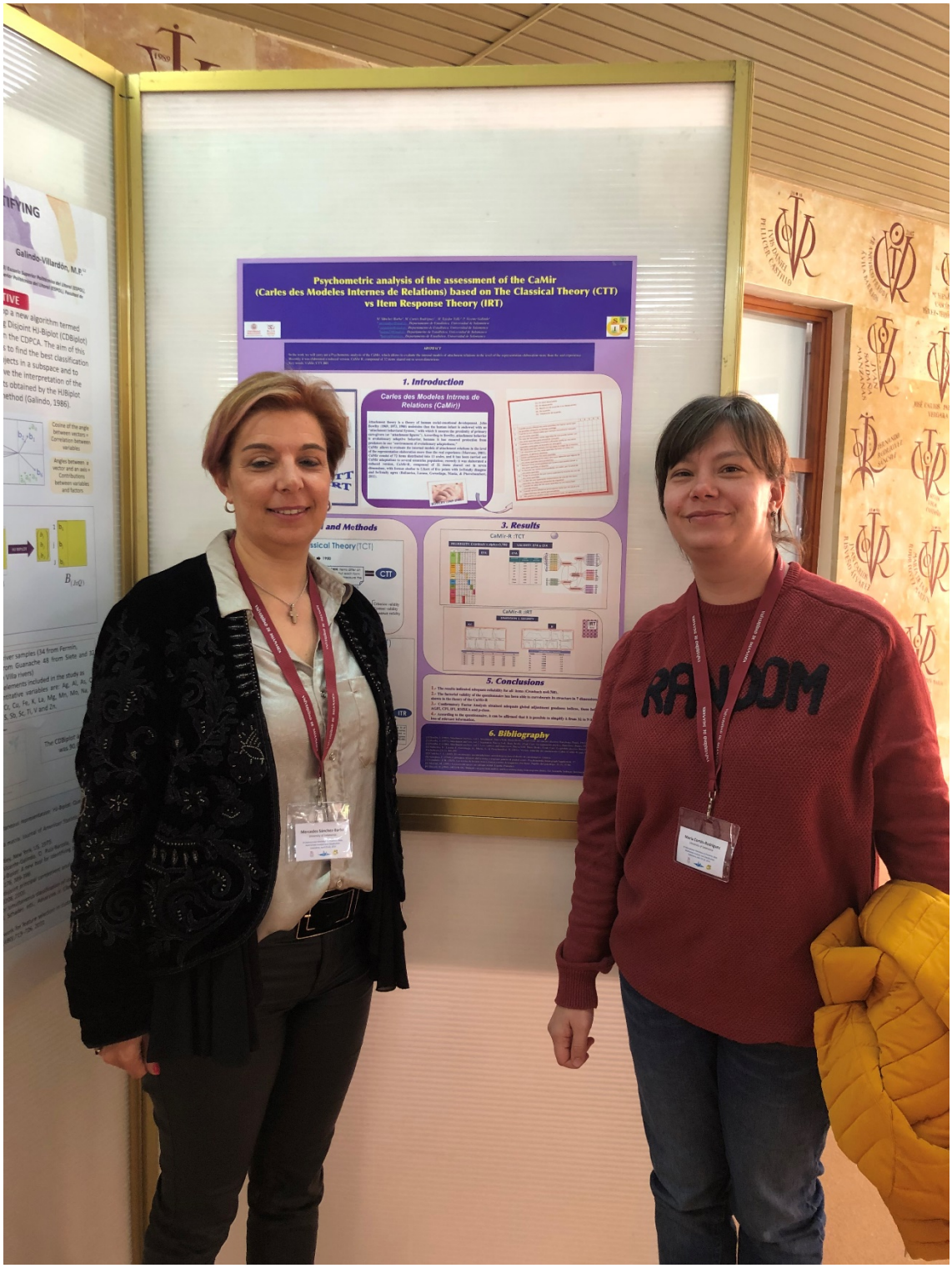

IV International Workshop on Proximity Data, Multivariate Analysis and Classification
WATER QUALITY MULTIVARIATE ANALYSIS: APPLICATION TO GATUN LAKE
ABSTRACT
INTRODUCTION
OBJECTIVE
MATERIAL AND METHODS
RESULTS

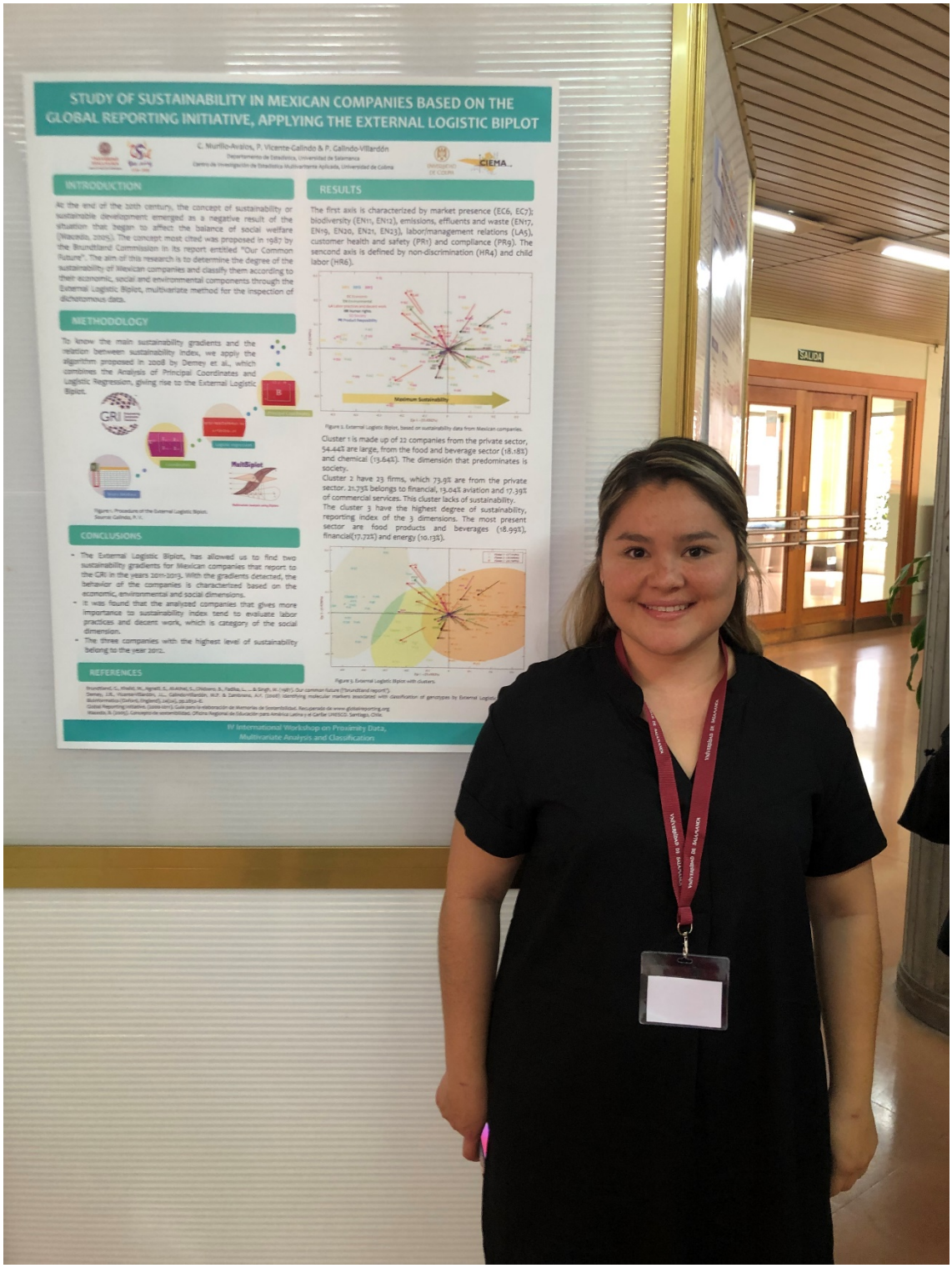

HIERARCHICAL MODELING AND COPULAS FOR THE STUDY OF DIAGNOSTIC ACCURACY
PROBLEM
CONTRIBUTIONS
METHOD
RESULTS
STATISTICAL MODELLING
HSROC Model, 2001
Bivariate Model, 2005
Copula Model, 2016
CONCLUSION AND REMARKS
SOURCE CODE

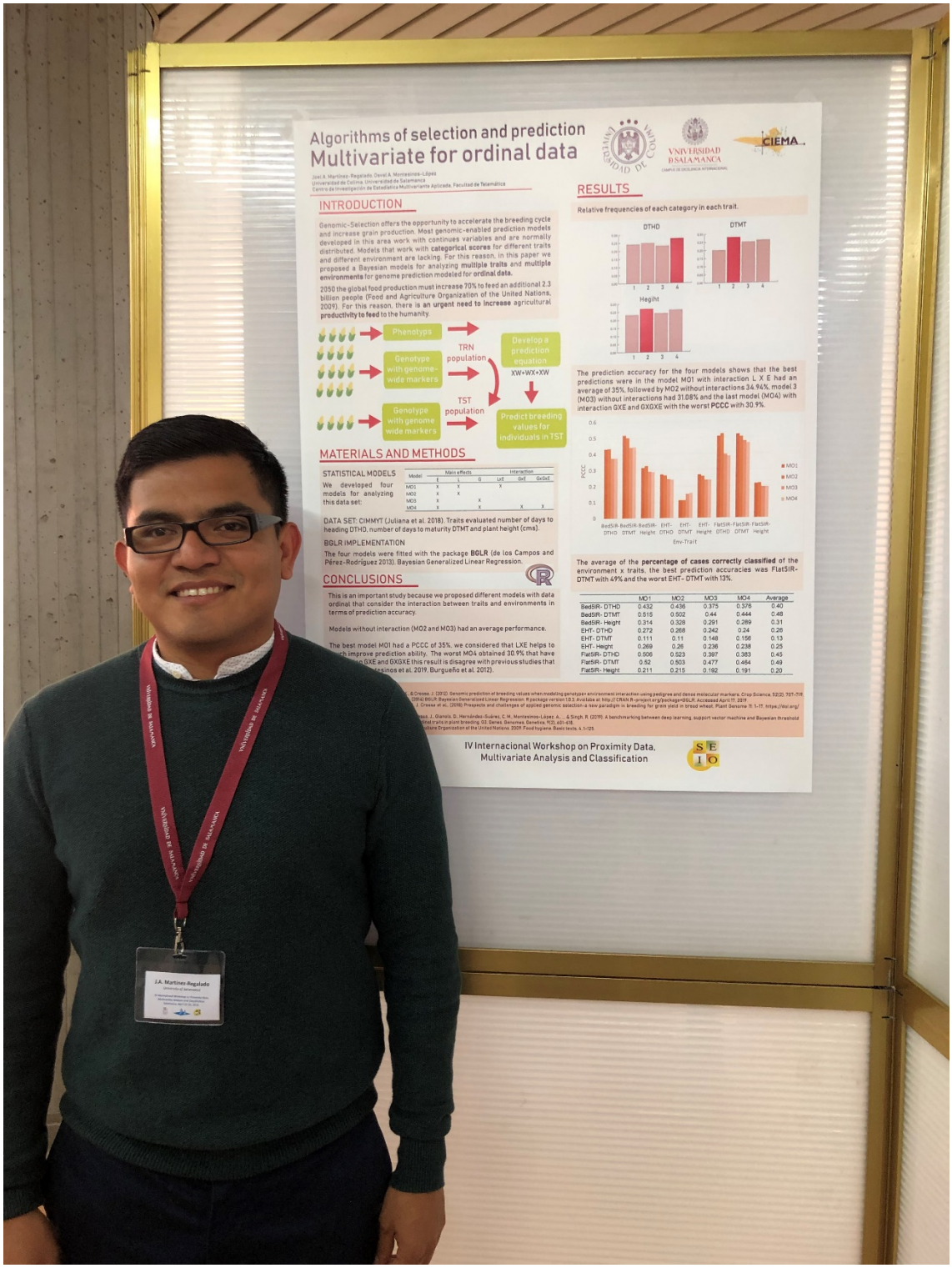

IV International Workshop on Proximity Data, Multivariate Analysis and Classification
Multivariate Characterization of Students of the University of Salamanca
Zárate-Santana, Z.J., Patino-Alonso, M.C., Sánchez-García, A.B., Celestino-Sánchez, M.A.
Universidad de Colima, Universidad de Salamanca
Introduction
In recent years numerous studies have contributed to boost the research on learning approaches and academic-stress coping by university students, evincing that the stress makes them change the way in which they learn. On the one hand, each student has a predetermined manner of approaching his/her learning. It can be superficial or deep, according to the degree of the involvement of students in their study determined by their creativeness, work capability and initiative to gather information beyond what professors provide (Hernández-Pina, García-Sanz, & Maquilón-Sánchez, 2005). On the other hand, students can cope with stress in three different fashions: positive re-evaluation (also known as cognitive reappraisal), searching for social support, and planning. The positive re-evaluation covers the strategy of focusing on the positive aspects and ignoring the negative ones of the problem. Students can also decide to seek social support (e.g., in lecturers, colleagues and/or family) in order to improve their emotional situation. Finally, planning consists of devising all the strategies to comprehend the problem and obtain an action plan to control the stress situation (Cabanach, Valle, Rodríguez, Piñeiro, & Freire, 2010).
Objetive
Determine the relationship between the learning approaches and the academic-stress coping of the students of the University of Salamanca.
Methods
The Canonical Correspondence Analysis (CCA) through the programme CANOCO is used to relate the composition of the student community with the known variation in the university environment (Ter Braak, 1986).
Sample
1,012 students of several knowledge areas of the University of Salamanca participated in the questionnaires.
Instruments
Two questionnaires are employed. First, the questionnaire R-SPQ-2F, which is composed of 20 items with Likert-scale-like answers, is designed to measure the learning approach of students in their current teaching context. It is therefore an instrument to evaluate teaching rather than characterising students as "surface learners" or "deep learners" (Biggs, Kember, & Leung, 2001). Second, the academic-stress coping is quantified by the questionnaire A-CEA, which is a subscale of the questionnaire CEA (Spanish initials for "Questionnaire for Academic Stress") consisting of 23 items. It is formulated to evaluate the cognitive strategies and behaviours that allow the students to face the situations of academic stress, discerning between cognitive reappraisal, searching for social support and planning (Cabanach et al., 2010).
Questionnaire R-SPQ-2F
DIMENSION | ITEMS
Deep approach | 1,2,5,6,9,10,13,14,17,18
Surface approach | 3,4,7,8,11,12,15,16,19,20
Questionnaire A-CEA
DIMENSION | ITEMS
Cognitive Reappraisal | 1, 4, 7, 10, 13, 16, 18, 20, 22
Searching for social | 2, 5, 8, 11, 14, 17, 21
Planning | 3, 6, 9, 12, 15, 19, 23
Results
The dimensions of the questionnaires are confirmed in our sample and we find a correlation between learning approaches and academic-stress coping. Results revealed that, when students are under academic stress, social-support searching is preferred by surface learners while planning is the favourite strategy of deep learners.
Figure 1. CCA of learning approaches and academic-stress coping.
Conclusions
We find a correlation between learning approaches and academic-stress coping.
References
Biggs, J., Kember, D., & Leung, D. Y. P. (2001). The Revised Two Factor Study Process Questionnaire: R-SPQ-2F. 71, 133–149.
Cabanach, R. G., Valle, A., Rodríguez, S., Piñeiro, I., & Freire, C. (2010). Escala de afrontamiento del estrés académico (A-CEA). Revista Iberoamericana de Psicología y Salud, 1(1), 51–64.
Hernández-Pina, F., García-Sanz, M. P., & Maquilón-Sánchez, J. (2005). Análisis del cuestionario de procesos de estudio-2 factores de Biggs en Estudiantes Universitarios Españoles. Revista Fuentes, 6.
Ter Braak, C. J. F. (1986). Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis. Ecology, 67(5), 1167–1179. https://doi.org/10.2307/1938672

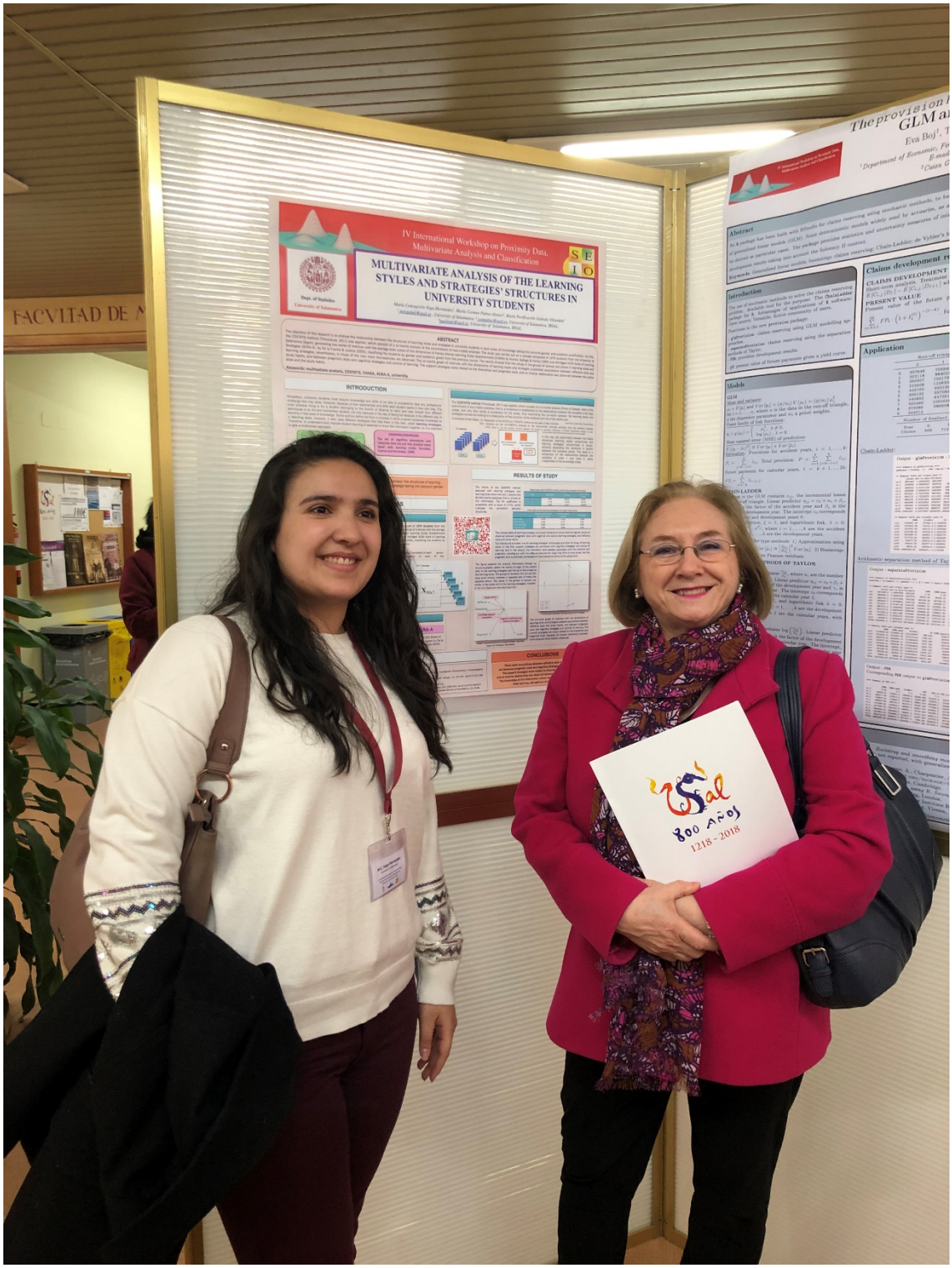

IV International Workshop on Proximity Data, Multivariate Analysis and Classification
SEIO
MULTIVARIATE ANALYSIS OF THE LEARNING STYLES AND STRATEGIES' STRUCTURES IN UNIVERSITY STUDENTS
Dept. of Statistics
University of Salamanca
ABSTRACT
INTRODUCTION
OBJECTIVE
MATERIAL AND METHODS
Participants
Instruments
CHAEA
ACRA-A
Analysis
RESULTS OF STUDY
CONCLUSIONS
REFERENCES

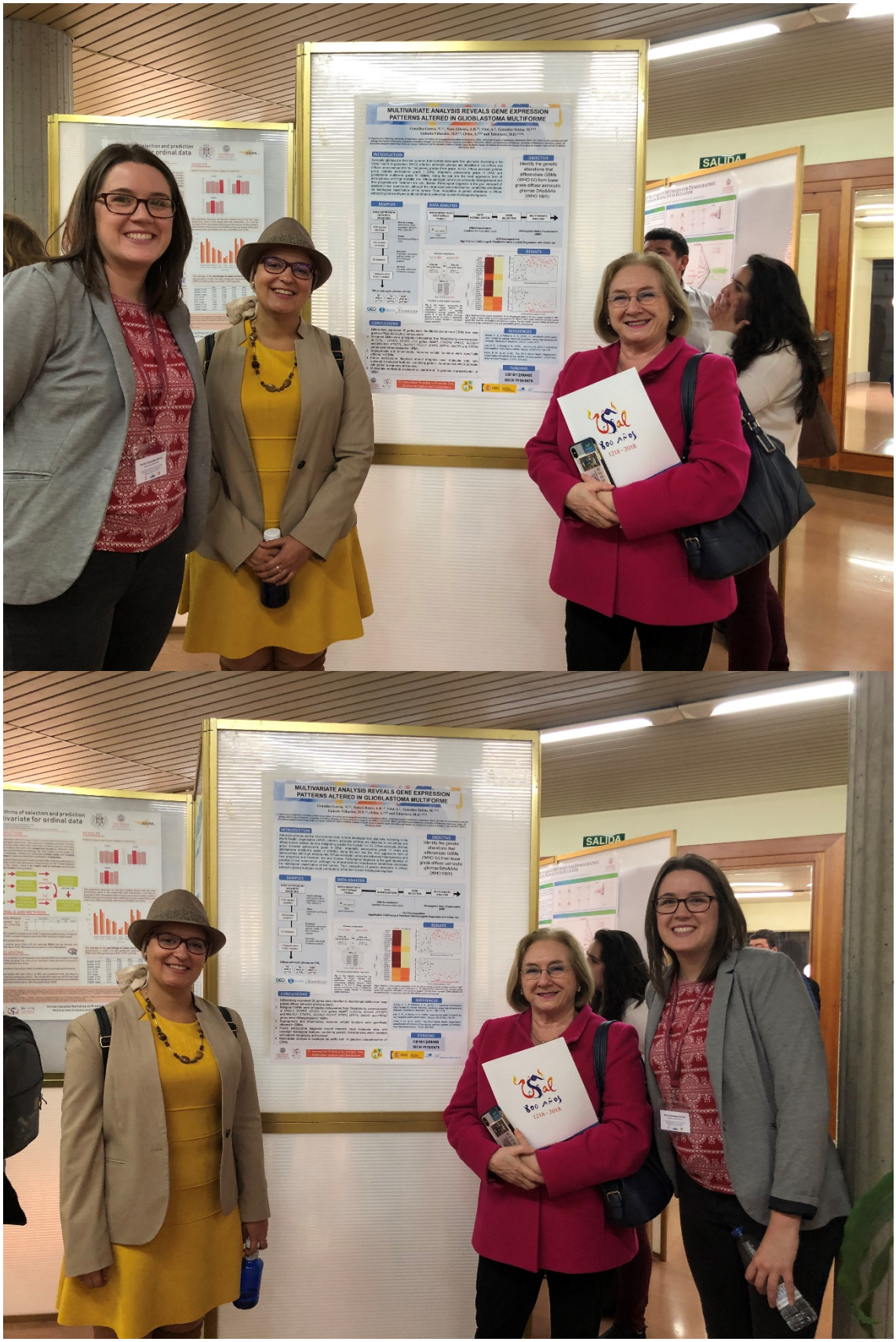

APPLICATION OF THE RESPONSE THEORY TO THE ITEM OF THE PSYCHOLOGICAL WELL-BEING QUESTIONNAIRE OF CAROL RYFF.
INTRODUCTION
METHODOLOGY
RESULTS
DISCUSSION & CONCLUSIONS

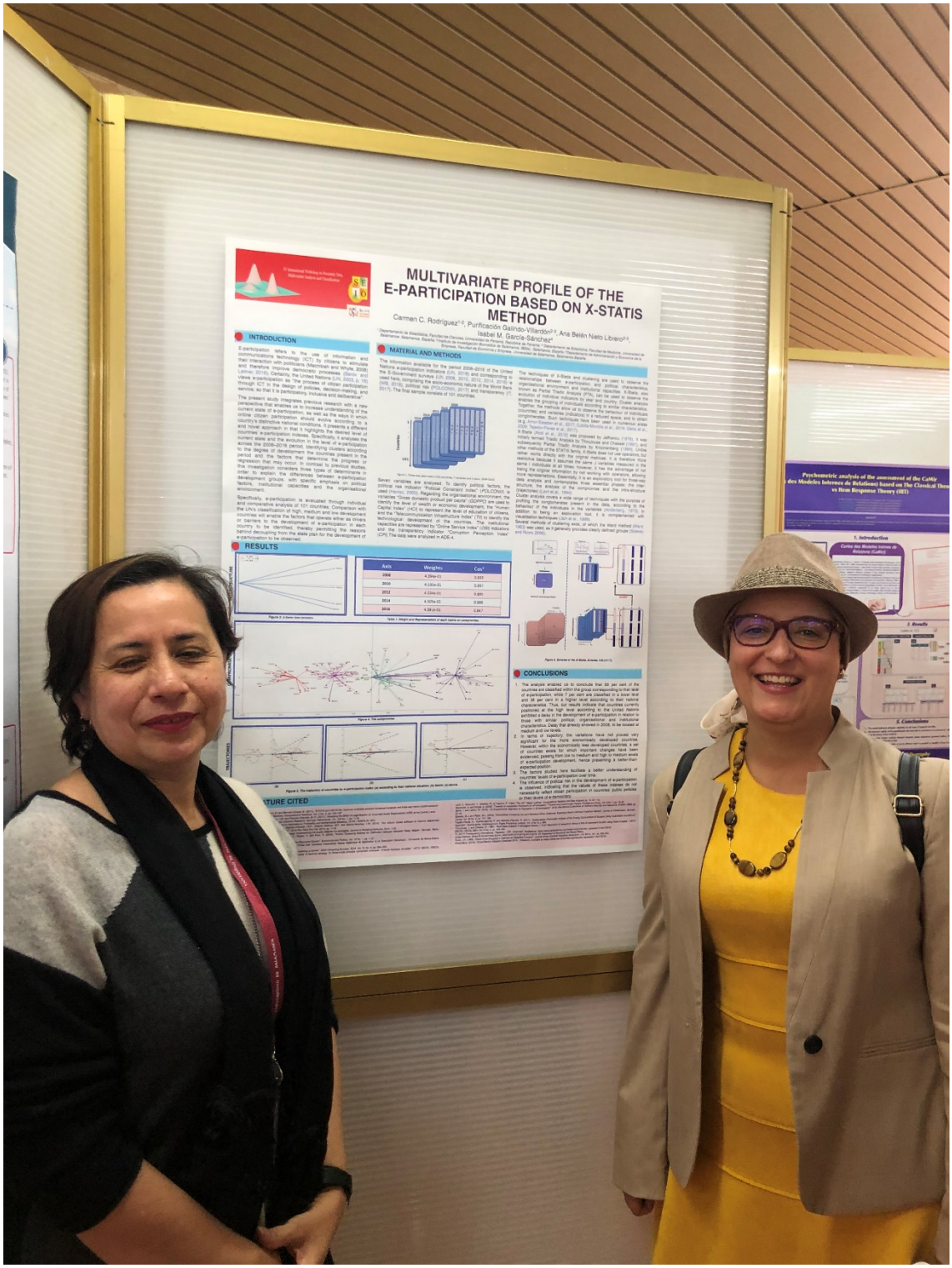

Psychometric analysis of the assessment of the CaMir
(Carles des Modeles Internes de Relations) based on The Classical Theory (CTT)
vs Item Response Theory (IRT)
1. Introduction
Carles des Modeles Internes de Relations (CaMir)
3. Results
5. Conclusions
6. Bibliography
RANDOM

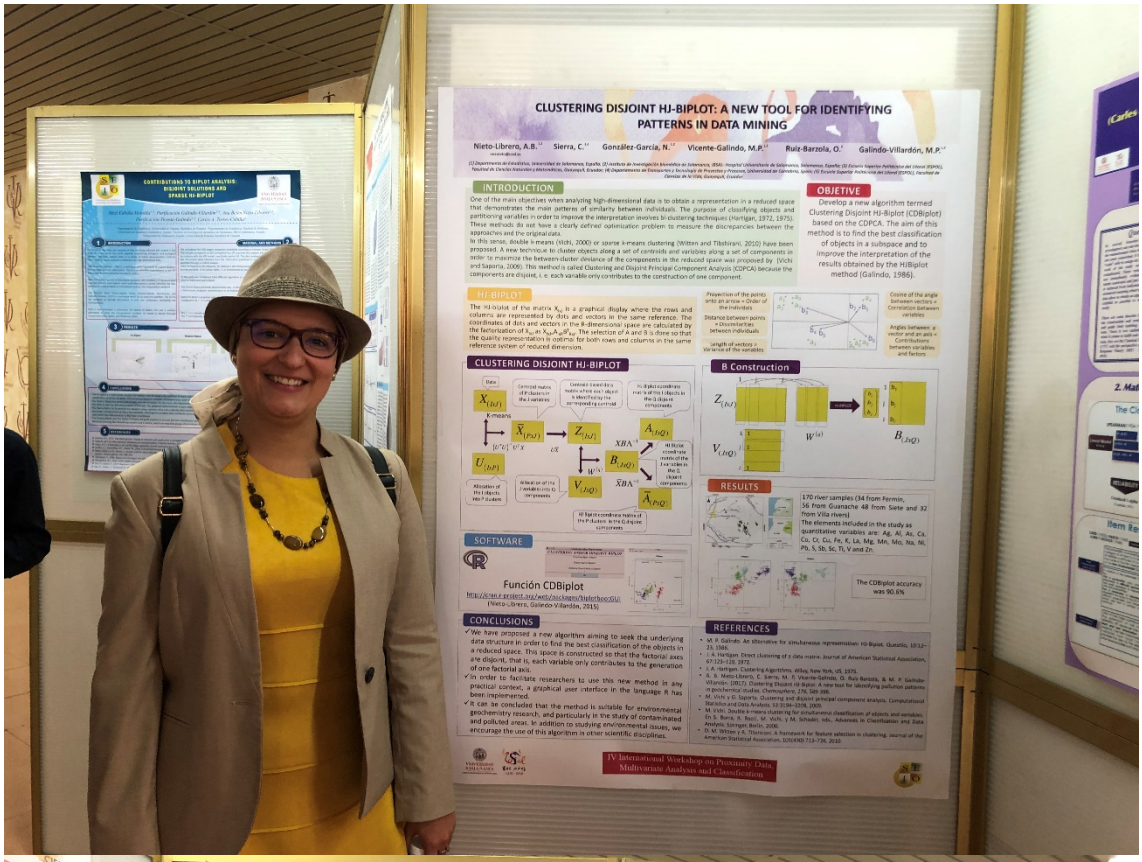
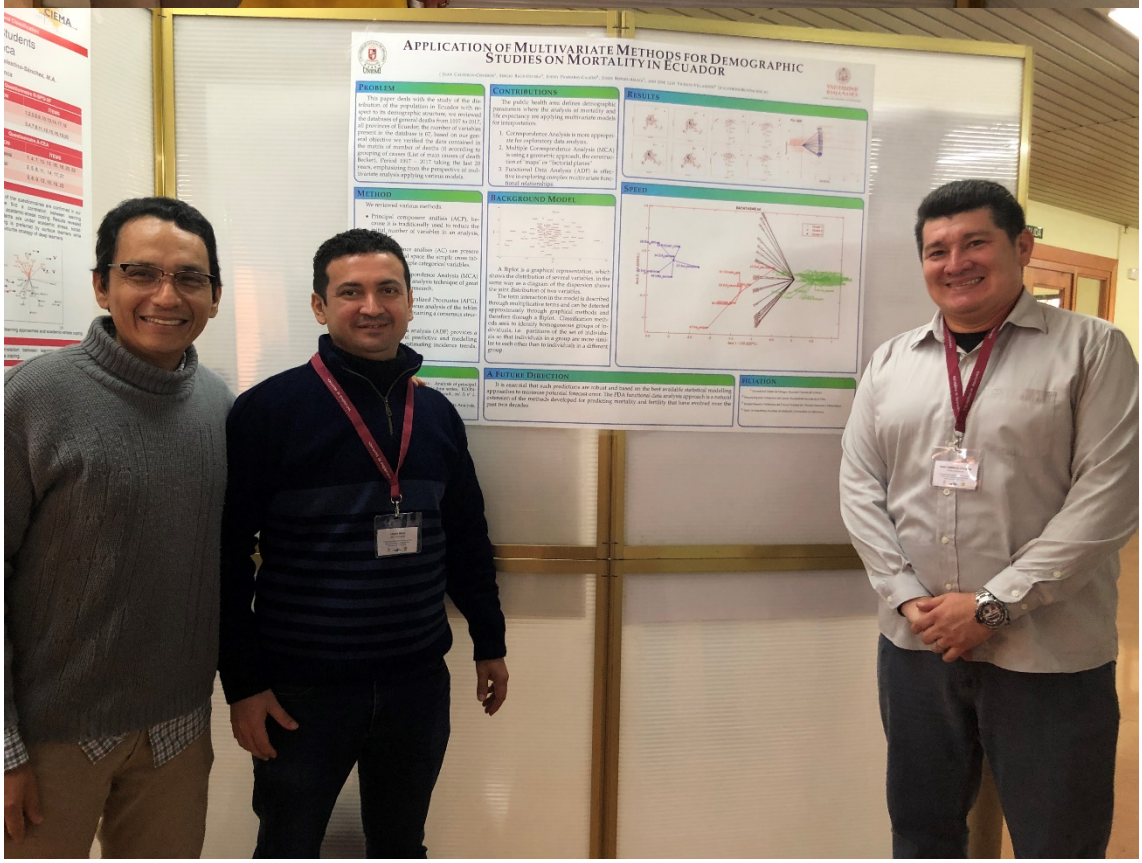

STUDY OF SUSTAINABILITY IN MEXICAN COMPANIES BASED ON THE GLOBAL REPORTING INITIATIVE, APPLYING THE EXTERNAL LOGISTIC BIPLOT
C. Murillo-Avalos, P. Vicente-Galindo & P. Galindo-Villardón
Departamento de Estadística, Universidad de Salamanca
Centro de Investigación de Estadística Multivariante Aplicada, Universidad de Colima
CIEMA
INTRODUCTION
At the end of the 20th century, the concept of sustainability or sustainable development emerged as a negative result of the situation that began to affect the balance of social welfare (Wacedo, 2005). The concept most cited was proposed in 1987 by the Brundtland Commission in its report entitled "Our Common Future". The aim of this research is to determine the degree of the sustainability of Mexican companies and classify them according to their economic, social and environmental components through the External Logistic Biplot, multivariate method for the inspection of dichotomous data.
METHODOLOGY
To know the main sustainability gradients and the relation between sustainability index, we apply the algorithm proposed in 2008 by Demey et al., which combines the Analysis of Principal Coordinates and Logistic Regression, giving rise to the External Logistic Biplot.
GRI
MultiBiplot
CONCLUSIONS
The External Logistic Biplot, has allowed us to find two sustainability gradients for Mexican companies that report to the GRI in the years 2011-2013. With the gradients detected, the behavior of the companies is characterized based on the economic, environmental and social dimensions.
It was found that the analyzed companies that gives more importance to sustainability index tend to evaluate labor practices and decent work, which is category of the social dimension.
The three companies with the highest level of sustainability belong to the year 2012.
REFERENCES
RESULTS
The first axis is characterized by market presence (EC6, EC7); biodiversity (EN11, EN12), emissions, effluents and waste (EN17, EN19, EN20, EN21, EN23), labor/management relations (LA5), customer health and safety (PR1) and compliance (PR9). The second axis is defined by non-discrimination (HR4) and child labor (HR6).
Minimum Sustainability
Cluster 1 is made up of 22 companies from the private sector, 54.44% are large, from the food and beverage sector (18.18%) and chemical (13.64%). The dimensión that predominates is society.
Cluster 2 have 23 firms, which 73.9% are from the private sector. 21.73% belongs to financial, 13.04% aviation and 17.39% of commercial services. This cluster lacks of sustainability.
The cluster 3 have the highest degree of sustainability, reporting index of the 3 dimensions. The most present sector are food products and beverages (18.99%), financial(17.72%) and energy (10.13%).
IV International Workshop on Proximity Data, Multivariate Analysis and Classification
SALIDA

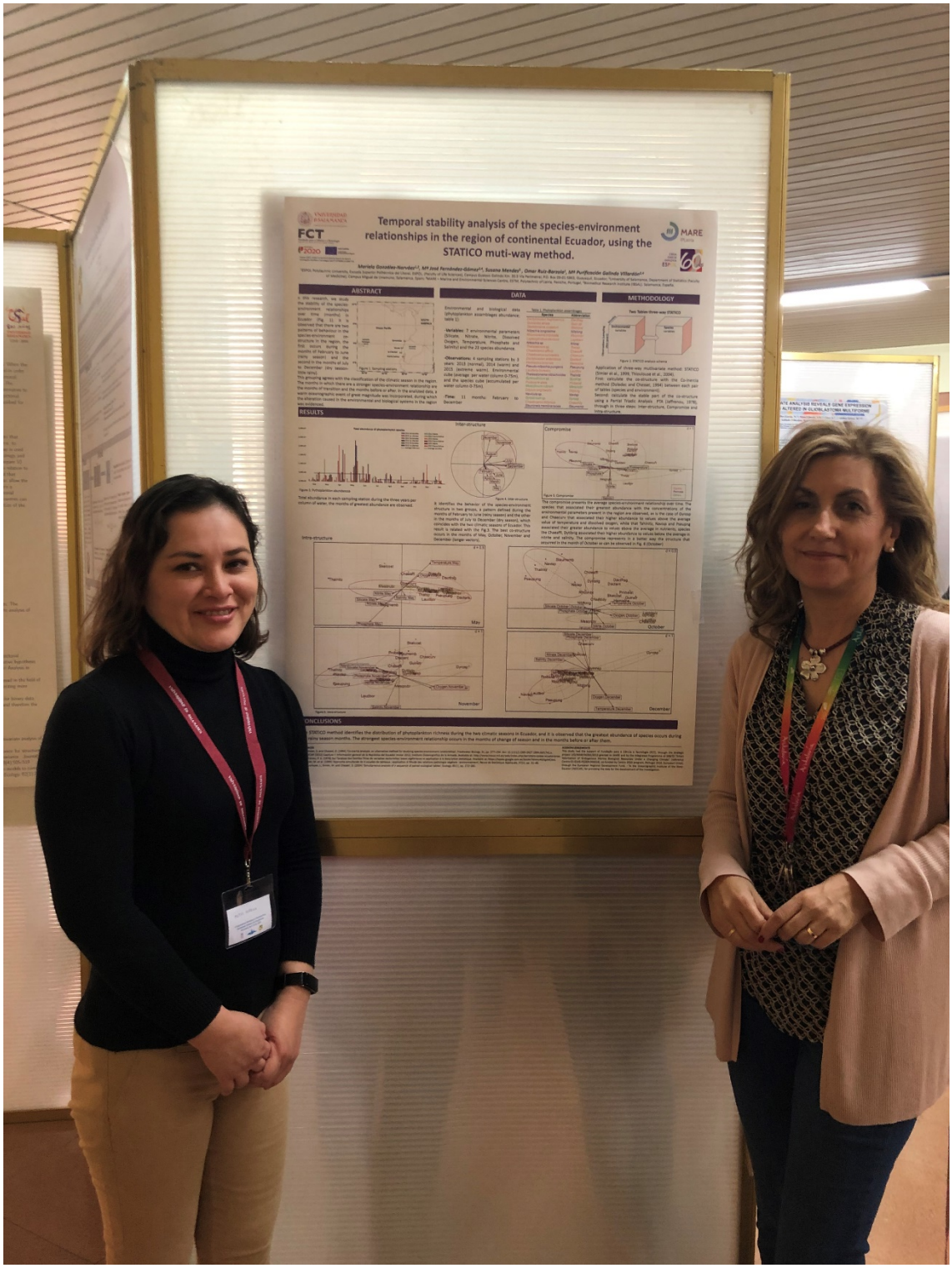

Algorithms of selection and prediction
Multivariate for ordinal data
INTRODUCTION
Genomic-Selection offers the opportunity to accelerate the breeding cycle and increase grain production. Most genomic-enabled prediction models developed in this area work with continues variables and are normally distributed. Models that work with categorical scores for different traits and different environment are lacking. For this reason, in this paper we proposed a Bayesian models for analyzing multiple traits and multiple environments for genome prediction modeled for ordinal data.
2050 the global food production must increase 70% to feed an additional 2.3 billion people (Food and Agriculture Organization of the United Nations, 2009). For this reason, there is an urgent need to increase agricultural productivity to feed to the humanity.
Phenotyps
Genotype with genome-wide markers
Genotype with genome wide markers
TRN population
TST population
Develop a prediction equation
XW+WX+XW
Predict breeding values for individuals in TST
MATERIALS AND METHODS
STATISTICAL MODELS
We developed four models for analyzing this data set:
DATA SET: CIMMYT (Juliana et al. 2018). Traits evaluated number of days to heading DTHD, number of days to maturity DTMT and plant height (cms).
BGLR IMPLEMENTATION
The four models were fitted with the package BGLR (de los Campos and Pérez-Rodríguez 2013). Bayesian Generalized Linear Regression.
CONCLUSIONS
This is an important study because we proposed different models with data ordinal that consider the interaction between traits and environments in terms of prediction accuracy.
Models without interaction (MO2 and MO3) had an average performance.
The best model MO1 had a PCCC of 35%, we considered that LXE helps to improve prediction ability. The worst MO4 obtained 30.9% that have GXE and GXGXE this result is disagree with previous studies that
RESULTS
Relative frequencies of each category in each trait.
DTHD
DTMT
Height
The prediction accuracy for the four models shows that the best predictions were in the model MO1 with interaction L X E had an average of 35%, followed by MO2 without interactions 34.94%, model 3 (MO3) without interactions had 31.08% and the last model (MO4) with interaction GXE and GXGXE with the worst PCCC with 30.9%.
PCCC
Env-Trait
MO1
MO2
MO3
MO4
The average of the percentage of cases correctly classified of the environment x traits, the best prediction accuracies was FlatSIR-DTMT with 49% and the worst EHT- DTMT with 13%.
MO1 MO2 MO3 MO4 Average
BedSIR- DTHD 0.432 0.436 0.375 0.378 0.40
BedSIR- DTMT 0.515 0.502 0.44 0.444 0.48
BedSIR- Height 0.314 0.328 0.291 0.289 0.31
EHT- DTHD 0.272 0.268 0.242 0.24 0.26
EHT- DTMT 0.111 0.11 0.148 0.156 0.13
EHT- Height 0.269 0.26 0.236 0.238 0.25
FlatSIR- DTHD 0.506 0.523 0.397 0.383 0.45
FlatSIR- DTMT 0.52 0.503 0.477 0.464 0.49
FlatSIR- Height 0.211 0.215 0.192 0.191 0.20
IV Internacional Workshop on Proximity Data, Multivariate Analysis and Classification
J.A. Martínez-Regalado

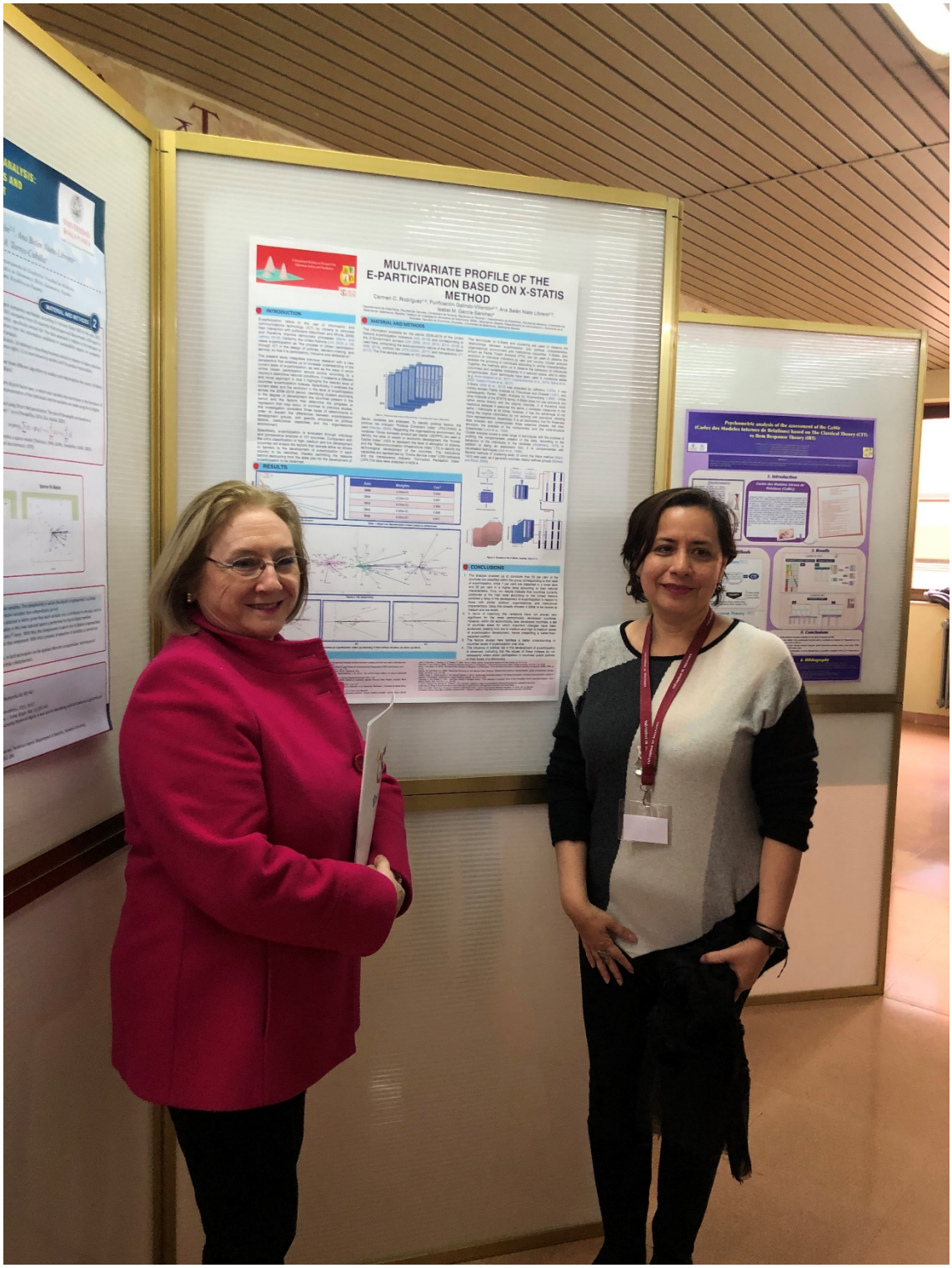

IV International Workshop on Proximity Data, Multivariate Analysis and Classification
MULTIVARIATE ANALYSIS OF THE LEARNING STYLES AND STRATEGIES' STRUCTURES IN UNIVERSITY STUDENTS
ABSTRACT
INTRODUCTION
RESULTS OF STUDY
CONCLUSIONS
800 AÑOS
1218 - 2018

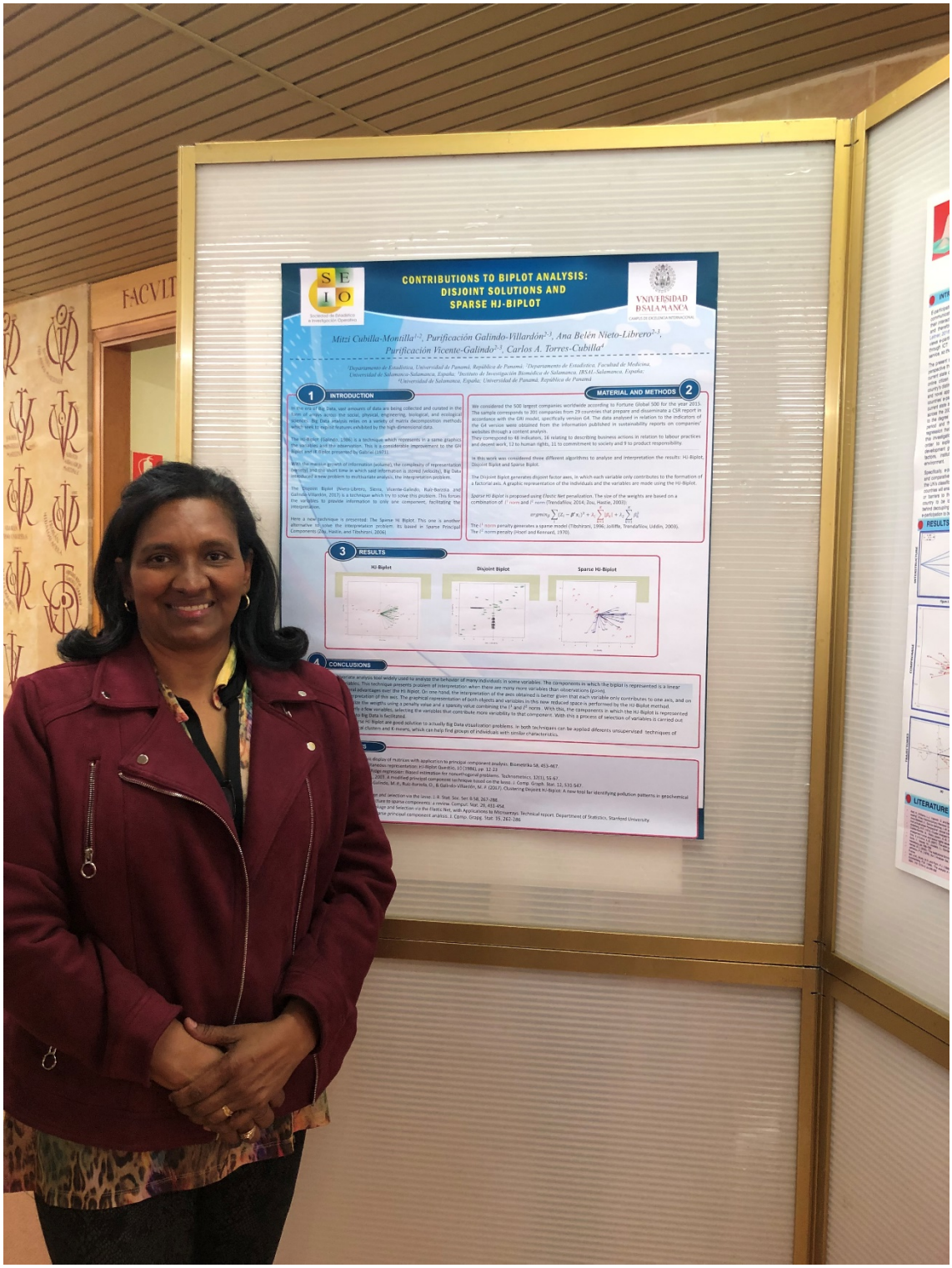

MULTIVARIATE ANALYSIS REVEALS GENE EXPRESSION PATTERNS ALTERED IN GLIOBLASTOMA MULTIFORME
SALIDA

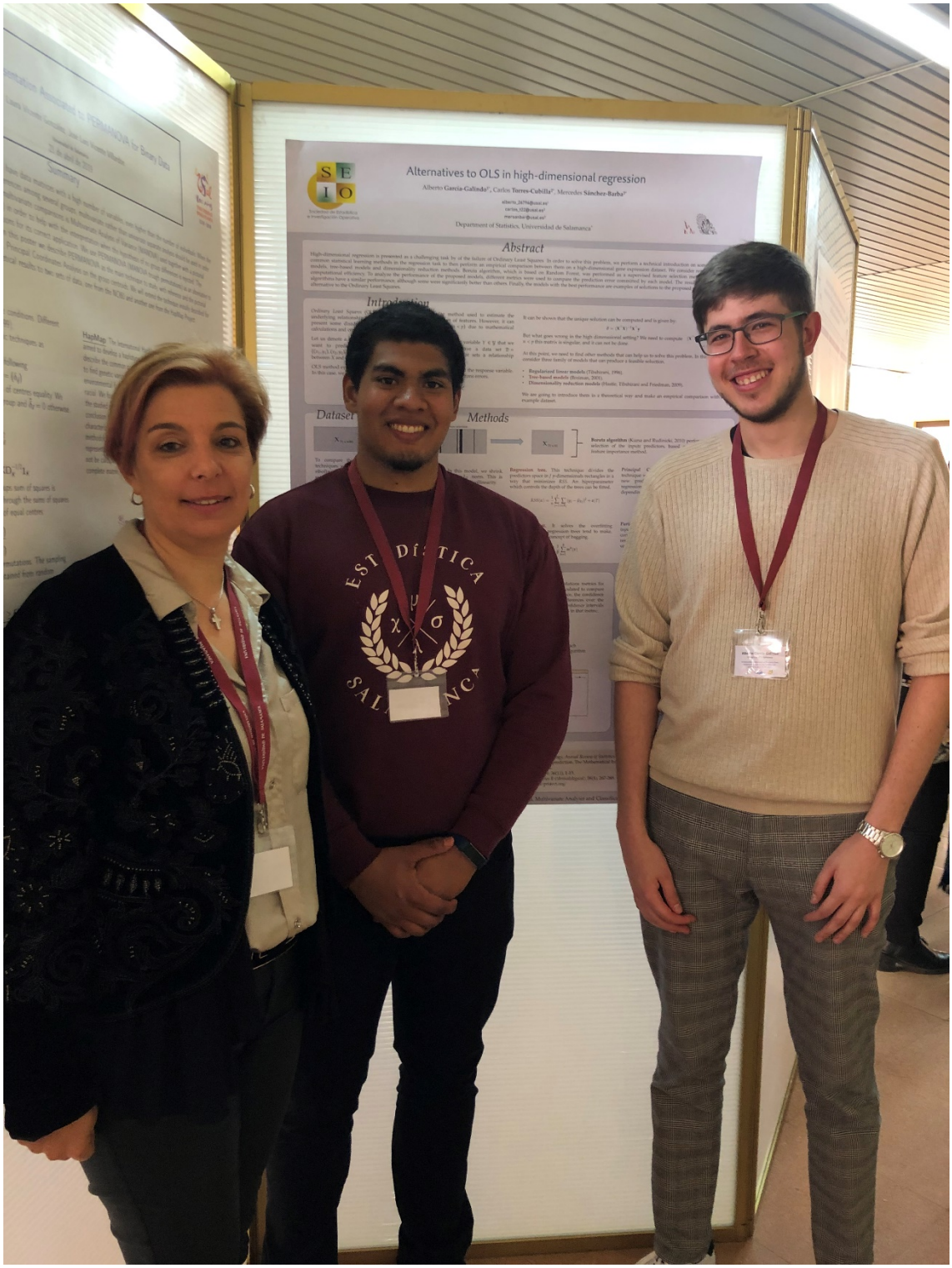

MULTIVARIATE PROFILE OF THE E-PARTICIPATION BASED ON X-STATIS METHOD
Carmen C. Rodríguez[1,2], Purificación Galindo-Villardón[2,3], Ana Belén Nieto Librero[2,3], Isabel M. García-Sánchez[4]
INTRODUCTION
MATERIAL AND METHODS
RESULTS
CONCLUSIONS
LITERATURE CITED


CLUSTERING DISJOINT HJ-BIPLOT: A NEW TOOL FOR IDENTIFYING PATTERNS IN DATA MINING
Nieto-Librero, A.B.
Sierra, C.
González-García, N.
Vicente-Galindo, M.P.
Ruiz-Barzola, O.
Galindo-Villardón, M.P.
INTRODUCTION
OBJETIVE
Develop a new algorithm termed Clustering Disjoint HJ-Biplot (CDBiplot) based on the CDPCA. The aim of this method is to find the best classification of objects in a subspace and to improve the interpretation of the results obtained by the HJBiplot method (Galindo, 1986).
HJ-BIPLOT
CLUSTERING DISJOINT HJ-BIPLOT
B Construction
RESULTS
SOFTWARE
Función CDBiplot
CONCLUSIONS
REFERENCES
IV International Workshop on Proximity Data Multivariate Analysis and Classification

APPLICATION OF MULTIVARIATE METHODS FOR DEMOGRAPHIC STUDIES ON MORTALITY IN ECUADOR
PROBLEM
CONTRIBUTIONS
RESULTS
METHOD
BACKGROUND MODEL
SPEED
A FUTURE DIRECTION
RELATION


Temporal stability analysis of the species-environment relationships in the region of continental Ecuador, using the STATICO muti-way method.
ABSTRACT
DATA
METHODOLOGY
RESULTS
CONCLUSIONS


MULTIVARIATE PROFILE OF THE E-PARTICIPATION BASED ON X-STATIS METHOD
INTRODUCTION
MATERIAL AND METHODS
RESULTS
CONCLUSIONS


CONTRIBUTIONS TO BIPLOT ANALYSIS:
DISJOINT SOLUTIONS AND
SPARSE HJ-BIPLOT
Mitzi Cubilla-Montilla[1,2], Purificación Galindo-Villardón[2,3], Ana Belén Nieto-Librero[2,3],
Purificación Vicente-Galindo[2,3], Carlos A. Torres-Cubilla[4]
1 INTRODUCTION
MATERIAL AND METHODS 2
3 RESULTS
HJ-Biplot
Disjoint Biplot
Sparse HJ-Biplot
4 CONCLUSIONS
LITERATURE


Alternatives to OLS in high-dimensional regression
Abstract
Methods
ESTADÍSTICA
SALAMANCA


A Graphical Representation Associated to PERMANOVA for Binary Data
Summary
PERMANOVA
HapMap
Conclusions
References

IV International Workshop on Proximity Data, Multivariate Analysis and Classification

April 25-26, 2019

Salamanca (Spain)