Sociedad Española De Biometría

Región Española de la IBS- Spanish Region of the International Biometric Society

Primeras Jornadas Científicas de Estudiantes de la SEB

Valencia, 19 y 20 de enero de 2015

Programa completo

Lunes 19 de Enero de 2015

12:00 – 12:10 **Presentación de las jornadas:** David Conesa y Natàlia Vilor-Tejedor

12:10 – 13:30 **Sesión 1.** *Aplicaciones Biométricas.* Moderador: Irantzu Barrio

12:10 – 12:30 **Natàlia Vilor-Tejedor:** *Statistical methods in Imaging Genetics*
12:30 – 12:50 **Nerea Gutiérrez:** *La importancia de una correcta estimación del LC50 en estudios toxicológicos*
12:50 – 13:10 **Elena Lázaro:** *The health state of organic vegetables*
13:10 – 13:30 **Iosu Paradinas:** *Species distribution modeling. New fisheries management requirements need new modeling approaches*

13:30 – 15:00 Comida

15:00 – 17:00 **Curso:** *Introducción a la Inferencia Bayesiana;* Anabel Forte (UV).

17:00 – 18:00 **Sesión de Pósteres**; aula 1.7

18:00 – 18:30 Coffee Break

18:30 – 19:15 **Mesa Redonda:** *Qué ofrece la SEB a sus miembros estudiantes y viceversa.* Natàlia Vilor-Tejedor (moderador), David Conesa, Vicente Gallego y Anabel Forte.

20:30 Actividades Sociales

Martes 20 de Enero de 2015

9:30 – 10:50 **Sesión 2.** *Análisis de Datos Biométricos I.* Moderador: Natàlia Vilor-Tejedor

09:30 – 09:50 **Irantzu Barrio:** *About the categorization of continuous variables in prediction models: method development and implementation*
09:50 – 10:10 **Manuel Higueras:** *An R package for the Generalized Hermite distributions*
10:10 – 10:30 **Moisés Gómez:** *Selecting the primary endpoint in a randomized clinical trial with CompARE. A cardiovascular case study*
10:30 – 10:50 **Martí Casals:** *Parameter estimation of Poisson generalized linear mixed models based on three different statistical principles: a Simulation Study*

10:50 – 11:20 Coffee Break

11:20 – 13:20 **Sesión 3.** *Análisis de Datos Biométricos II.* Moderador: Manuel Higueras

11:20 – 11:40 **Urko Aguirre:** *Assessment of the performance of imputation techniques in observational studies with two measurements*
11:40 – 12:00 **Karen Flórez:** *Un modelo de conglomerados con estructura de clases latentes para el análisis de mapas de enfermedades*
12:20 – 12:40 **Amanda Fernández:** *Estudio de la mortalidad bovina registrada en granja mediante modelos INAR, potencial como componente de vigilancia sindrómica*
12:40 – 13:00 **Ana Vázquez:** *Models for binary response and survival data*
13:00 – 13:20 **Diego Ayma:** *Spatial Disaggregation in Epidemiology*

13:20 – 13:45 **Presentación Congreso de Bilbao:** Irantzu Barrio

13:45 – 14:00 **Cierre de la Jornada:** David Conesa

Comité científico/organizador:

David Conesa (UV)
Irantzu Barrio (UPV/EHU)
Natàlia Vilor-Tejedor (CREAL)
María Oliveira (USC, ABANCA)
Vicente Gallego (UVIC)
Manuel Higueras (PHE, UAB)

Sesión 1

Aplicaciones Biométricas

Statistical methods in Imaging Genetics

Natàlia Vilor-Tejedor[1,2,3], Jordi Sunyer[1,2,3,4], Juan R. González[1,2,3]

(1) *Centre for Research in Environmental Epidemiology (CREAL), Barcelona, Spain.*

(2) *Universitat Pompeu Fabra (UPF), Barcelona, Spain*.

(3) *CIBER Epidemiology and Public Health (CIBERESP), Spain.*

(4) *IMIM (Hospital del Mar Medical Research Institute), Barcelona, Spain.*

Recent advances in biotechnology allow the development of new tools and techniques for biomedical investigations. This is especially relevant because there is a need to combine data from different sources using different methodologies. For instance, brain structure and function can be objectively studied from Magnetic Resonance Imaging techniques.

This type of data is commonly used to further confirm and reinforce neuropsychological evidences and interpretations in epidemiological studies. In this regard, environmental epidemiology can greatly benefit from neuroimaging studies examining the impact of potential environmental exposures to the developing brain structure and functioning.

Furthermore, more accurate findings can be obtained using genomics. The integration of genomic and neuroimaging data, a field known as *Imaging Genetics*, represents a challenge for current biomedical research. As a consequence, the development of novel statistical and mathematical approaches which efficiently analyze multimodal data is crucial to obtain relevant and reliable results.

The main goal of this proposal is to review some Imaging genetic strategies to assess potential associations between genetic biomarkers and brain development.

References:

1. Liu J, Calhoun VD. *A review of multivariate analyses in imaging genetics*. Front Neuroinform. 2014 Mar 26;8:29. PubMed PMID: 24723883.

2. Bookheimer SY, Strojwas MH, Cohen MS, Saunders AM, Pericak-Vance MA, Mazziotta JC, Small GW. *Patterns of brain activation in people at risk for Alzheimer's disease*. N Engl J Med. 2000 Aug 17;343(7): 450-6. PubMed PMID: 10944562.

3. Roussotte FF, Gutman BA, Hibar DP, Jahanshad N, Madsen SK, Jack CR Jr, Weiner MW, Thompson PM; Alzheimer's Disease Neuroimaging Initiative (ADNI). *A single nucleotide polymorphism associated with reduced alcohol intake in the RASGRF2 gene predicts larger cortical volumes but faster longitudinal ventricular expansion in the elderly*. Front Aging Neurosci. 2013 Dec 19;5:93. PubMed PMID: 24409144.

La importancia de una correcta estimación del LC50 en estudios toxicológicos

Nerea Gutierrez[1], Irantzu Barrio[1], José María Lacave[2], Amaia Orbea[2], Inmaculada Arostegui1[1]

[1] Departamento de Matemática Aplicada, Estadística e I.O. Universidad del País Vasco UPV/EHU
[2] Departamento de Zoología y Biología Celular Animal y Centro de Investigación en Biología y Biotecnología Marinas Experimentales-PIE. Universidad del País Vasco UPV/EHU

En el ámbito de la toxicología suele ser habitual trabajar con el valor LC50, la concentración de una sustancia que provoca la muerte del 50% de un grupo de organismos de prueba en un tiempo determinado, ya que este valor nos permite evaluar la toxicidad de las sustancias de una forma sistemática y comparable.

Recientemente, en un estudio sobre el efecto de las nanopartículas de plata en el desarrollo de los embriones de pez cebra del Grupo de Investigación de Biología Celular en Toxicología Ambiental de la UPV/EHU, observamos que la estimación del valor LC50 dependía del método de estimación y software utilizados, y que ante situaciones extremas, se observaba una separación de datos que provocaba que las estimaciones hechas con el método de máxima verosimilitud no fuesen correctas.

El objetivo de este estudio es comparar diferentes estrategias de estimación del LC50 y su implementación en diferentes software estadísticos. Supongamos que tenemos una variable respuesta dicotómica Y y X una variable continua que representa el nivel de concentración al que han sido expuestos los sujetos. El modelo que utilizaremos para explicar Y en función de X, será el modelo lineal generalizado tal que $g(\mu) = \beta_{0+}\beta_1 X$, siendo g la función link y $\mu = E(Y|X)$. En este caso se han comparado dos métodos de estimación de parámetros: máxima verosimilitud y máxima verosimilitud penalizada, propuesta por Firth (Heinze, 2006). En los casos en los que se da una "separación o quasi separación" de datos, el método de Firth proporciona estimaciones finitas de los parámetros, no así el de máxima verosimilitud. La implementación de estos métodos se ha comparado en los software SPSS y R. Hemos llevado a cabo un estudio de simulación en el que se muestra que la estimación del LC50 depende del método de estimación utilizado.

Referencias:

1. Heinze G. and Schemper M. (2002)
   A solution to the problem of separation in logistic regression. Statistics in Medicine, 21(1),2409 – 2419.

2. Heinze G. (2006)
   A comparative investigation of methods for logistic regression with separated or nearly separated data. Statistics in medicine, 25(1),4216 – 4226.

3. Firth D. (1993)
   Bias Reduction of maximun likelihood estimates. Biometrika, 80,27-38.

The health state of organic vegetables

E. Lázaro[1], *C. Armero*[1], *J. Roselló*[2], *J. Serra*[2], *M.J. Muñoz*[3], *L. Rubio*[4]

[1] Departament d'Estadística i Investigació Operativa. Universitat de València.
[2]Estació Experimental Agraria de Carcaixent. Instituto Valenciano de Investigaciones Agrarias.
[3]Laboratorio del Servicio de Sanidad Vegetal y Protección Fitosanitaria. Generalitat Valenciana.
[4]Centro de Protección Vegetal y Biotecnología. Instituto Valenciano de Investigaciones Agrarias.

Abstract

Vegetable production in Spain is mainly located in the Mediterranean area and generates a significant contribution to its economy. However, vegetable crops suffer heavy losses as a consequence of the damages caused by viruses. Serra *et al.* (1999) asserted that ecological and epidemiological factors responsible for most of the viral infections in the Spanish Mediterranean area are basically determined by the agroecosystem balance with a great variability in its intensity depending on the different growing conditions. Consequently, they stated that organic vegetable farming is a real alternative production system that minimizes the incidence of diseases caused by viruses while preserving the agroecosystem balance.

The aim of this study is to assess the proportion of plants infected with viruses with regard to the organic and non-organic production system as well as other relevant covariates which contain information about plot management. We analyzed two species of plants belonging to the *Solanacea* family, tomato and pepper, because of their wide distribution in the València region and three of the major *Solanaceus* endemic region viruses: *Cucumber mosaic virus*, *Tomato mosaic virus* and *Tomato spotted wilt virus* (Hansen and Lapidot, 2012).

A total of 30 plots were classified according to the production system (organic or non-organic) and 240 individual plants (8 randomly selected in each plot) were observed for symptoms and analyzed to determine the presence or absence of viral infections in each plant. Bayesian hierarchical logistic regression models have been considered to perform a marginal statistical analysis of the prevalence of each virus. A Bayesian logit generalized linear model for correlated binary data is considered for studying the joint distribution of the prevalence of the three viruses in order to capture possible interactions between them. Markov chain Monte Carlo methods have been used in order to approximate the posterior distribution of the parameters and hyperparameters of the model through the free software WinBUGS (Lunn *et al.*, 2000).

References

Lunn, D.J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS – A Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10: 325-337.

Serra, J., Ocon, C., Jiménez, A., Arnau, J., Malagón, J., y Porcuna, J. L. (1999). Epidemiología de las virosis en la Comunidad Valenciana: el caso del "virus de la cuchara" del tomate. *Comunidad Valenciana Agraria*, 14: 47-53.

Hanssen, I.M. and Lapidot, M. (2012). Major Tomato Viruses in the Mediterranean Basin. *Advances in Virus Research*, 84: 31-66.

Species distribution modeling. New fisheries management requirements need new modeling approaches.

Iosu Paradinas[1], Antonio Quiléz[1], Jose Mª Bellido[2] and David Conesa[1]

[1]Universitat de Valencia; [2]Instituto Español de Oceanografía

Las nuevas directrices de gestión pesquera comunitaria apuntan hacia un manejo espacial de las pesquerías (FAO 2008). Esto requiere una caracterización de los stocks y de sus interacciones en un marco espacial. Hasta ahora, la gran mayoría de enfoques y técnicas utilizadas en esta área no han alcanzado estos objetivos de manera cuantitativa. En cambio los últimos avances en geostadística Bayesiana, por medio del Integrated Nested Laplace Approximation (INLA) (Rue et al. 2009, Lindgren et al 2011), pueden permitir el ajuste de un amplio espectro de modelos interesantes para la gestión espacial pesquera en tiempos computacionales aceptables.

References:

1. Lindgren F, Rue H, Lindström J (2011) An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J Roy Stat Soc B Met 73:423-498
2. FAO (2008) The ecosystem approach to fisheries. FAO, Technical Guidelines for Responsible Fisheries
3. Rue H, Martino S, Chopin N (2009) Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J Roy Stat Soc B Met 71:319-392

Sesión 2

Análisis de Datos Biométricos I

About the categorization of continuous variables in prediction models: method development and implementation

Irantzu Barrio[1], Inmaculada Arostegui[1] and María-Xosé Rodríguez-Álvarez[2]

[1]Departamento de Matemática Aplicada, Estadística e Investigación Operativa. Universidad del País Vasco UPV/EHU

[2]Departamento de Estadística e Investigación Operativa. Universidade de Vigo

In the development of clinical prediction models it is common to use categorical variables as predictors, especially when those models aim to be applied in daily clinical practice to support clinicians at decision time.

Consider we have a dichotomous response variable *Y* and a continuous covariate *X* which we want to categorize. Previous work has been done on the categorization of continuous variables in the context of the logistic regression model. However, they do not allow to categorize a continuous variable in a multivariate setting. In this work, we propose a new approach to categorize *X* in *k+1* intervals (X_{catk}), in such a way that the best predictive logistic model is obtained for X_{catk}. Since the area under the ROC curve (AUC) is the most widely used discriminatory ability measure in logistic regression, we consider as the best predictive model that for which the AUC is maximized. Two alternative algorithms are proposed to search the optimal *k* cut points, respectively named *AddFor* and *Genetic*. With the AddFor we look for each cut-point at a time while the Genetic algorithm simultaneously looks for the vector of *k* cut points using Genetic Algorithms, the most widely known type of evolutionary algorithm.

We have implemented the proposed methods in the R package catpredi. This package provides the user with an easy to use tool to categorize continuous variables in prediction models. Providing the dichotomous response variable *Y*, the continuous covariate *X* and *k*, the number of cut-points, the user can choose which method to use to categorize *X*. If the method chosen is the *AddFor*, an optional parameter can be provided to determine the grid size. This library will return the k optimal cut-points, as well as the final models AUC and the categorized variable X_{catk} .

Financiación: 2012111008, IT620-13, MTM2013-40941-P, UFI11/52.

References:

1. Tsuruta H., Bax L.(2006) Polychotomization of continuous variables in regression models based on the overall C index. *BMC Medical Informatics and Decision Making*, 6,41.
2. Barrio I, Arostegui I, Quintana J, IRYSS-COPD-GROUP. Use of generalised additive models to categorise continuous variables in clinical prediction. *BMC Medical Research Methodology* 2013; 13:83.

An *R* package for the Generalized Hermite distributions

David Moriña[1], Manuel Higueras[2,3], Pedro Puig[2]

[1]CREAL; [2]UAB; [3]PHE

Generalized Hermite distributions are a family of two-parameter count distributions. These distributions can be useful for modelling count data that presents multi-modality or overdispersion (the variance greater than the mean), situations that appear commonly in practice in many fields. These distributions are closed under convolution and their maximum likelihood estimator of the population mean is the sample mean.

A Generalized Hermite distribution of order is represented by , where and are Poisson distributed independent random variables, and is an integer greater or equal to 2. The second order Generalized Hermite distribution is the classical Hermite distribution.

In this work we present a new *R* package that allows the user to work with the probability density, cumulative density, quantile and random generation functions (*d/p/q/r*) of the Generalized Hermite distributions.

When one (or both) of the population means of and is (are) greater than 20, the distribution function is approximated using an Edgeworth expansion, the probability mass function is calculated from this approximation of the distribution function, and the quantile function is approached by a Cornish-Fisher expansion.

The *hermite* package also allows the user to perform the likelihood ratio test for Poisson assumption and to estimate parameters using the maximum likelihood method.

Practical examples of the usage of these distributions can be found in economy and biology fields. One of them comes from a very recent publication that presents a Bayesian-type methodology for dose estimation in cytogenetic dosimetry.

References:

1. Moriña D, Higueras M, Puig P. The *R* Package *hermite* (submitted).
2. Puig P (2003). Characterizing Additively Closed Discrete Models by a Property of Their Maximum Likelihood Estimators, with an Application to Generalized Hermite Distributions. *Journal of the American Statistical Association*, **98**(463), 687-692.
3. Higueras M, Puig P, Ainsbury EA, Rothkamm K (2015). A new Inverse Regression Model Applied to Radiation Biodosimetry. *Proceedings of the Royal Society A.*

Selecting the primary endpoint in a randomized clinical trial with *CompARE*. A cardiovascular case study

Moisés Gómez-Mateu, Guadalupe Gómez

Universitat Politècnica de Catalunya

The appropriate choice of the primary endpoint is crucial at the design stage of randomized clinical trials. The decision is often in terms of whether or not secondary endpoints have to be added in order to detect the desired effect of the treatment under investigation.

Gómez and Lagakos (2013)[(1)] develop a methodology to evaluate the convenience of using a composite endpoint as the primary endpoint instead of one of its components. Their method evaluates the relative efficiency (ARE) of the composite endpoint versus a relevant subset of its components. The ARE can be interpreted as the ratio of the required sample sizes to detect a specific treatment effect for a given significance level and power[(2)].

In this talk, we present *CompARE*[(3)], a free web-based tool for trialists that computes the ARE for several combinations of components and sets the ground for a more efficient choice of the primary endpoint. The user can interact with *CompARE* through HTML form pages and no knowledge of R is needed.

In order to use *CompARE*, a list of candidate endpoints together with the anticipable probabilities and the relative treatment effect given by the hazard ratio have to be provided. The researcher can combine several components and obtain the ARE for every specific comparison. Results are shown immediately by means of plots and tables.

The choice of the primary endpoint in the cardiovascular area is of relevant importance and several clinical trials have failed to show significance of the primary endpoint due to their wrong choice. We illustrate some of the capabilities of this platform by means of a case study.

References:

1. Gómez G and Lagakos SW. (2013). Statistical considerations when using a composite endpoint for comparing treatment groups. Statistics in Medicine, 32, 19-738.
2. Gómez G, and Gómez-Mateu M. (2014). The Asymptotic Relative Efficiency and the ratio of sample sizes when testing two different null hypotheses. SORT, 38, 73-88.
3. Gómez-Mateu M and Gómez G. http://composite.upc.edu/CompARE (Last date of access: December 8, 2014).

Parameter estimation of Poisson generalized linear mixed models based on three different statistical principles: a Simulation Study

Martí Casals[1], Klaus Langohr[2], Josep Lluís Carrasco[3], Lars Rönnegard[4]

[1]marticasals@gmail.com, Bioestadística. Departament de Salut Pública. Universitat de Barcelona

[2]klaus.langohr@upc.edu,Department of Statistics and Operations Research, Universitat Politècnica de Catalunya- Barcelonatech, Spain

[3]jlcarrasco@ub.edu,Bioestadística. Departament de Salut Pública. Universitat de Barcelona

[4]lrn@du.se, Statistics Unit, Dalarna University, SE-791 88 Falun, Sweden

Abstract

Generalized linear mixed models (GLMMs) are a flexible approach to fit non-normal data. GLMMs are known to be useful for accommodating overdispersion in Poisson regression models and modeling the cluster dependence structure in longitudinal or repeated measures designs. The main difficulty of GLMMs is the parameter estimation because it is often not viable to obtain an analytic solution that allows maximizing the marginal likelihood of data. For this reason, a huge amount of attention has been paid in the statistical literature to this issue leading to several proposals that are able to estimate the parameters of a GLMM. Hence, it is possible to find different principles to fit GLMMs implemented in the main statistical software packages. Among them, we want to highlight Gauss-Hermite quadrature (GHQ) estimation, hierarchical (h-likelihood), or Bayesian estimation methods (Integrated Nested Laplace Approximation (INLA)) implemented in the R packages lme4, hglm, and INLA, respectively. The purpose of this study is to compare the performance of three different statistical principles {Marginal likelihood, Extended likelihood, Bayesian approach{ via a simulation study with different scenarios of overdispersion. A data example about injuries of contact wrestling is used for illustration. Advantages and limitations of these estimation methods are discussed in detail.

References:

1. Rue H, Martino S, Lindgren F, Simpson D, Riebler A, Krainski ET. Inla: Functions which allow to perform full bayesian analysis of Latent gaussian models using integrated nested laplace approximaxion. 2014; R package version 0.0-1404466487.
2. Bates D, Maechler M, Bolker B,Walker S. lme4: Linear mixed-e_ects models using Eigen and S4. 2014; R package version 1.1-7; Available from: http://CRAN.R-project.org/package=lme4.
3. Lars Ronnegard, Xia Shen and Moudud Alam (2010). hglm: A Package for Fitting Hierarchical Generalized Linear Models. *The R Journal*. URL: http://journal.r-project.org/archive/2010-2/RJournal_2010-2_Roennegaard~et~al.pdf.

Sesión 3

Análisis de Datos Biométricos II

Assessment of the performance of imputation techniques in observational studies with two measurements

Urko Aguirre[1], Inmaculada Arostegui[2], Jose M. Quintana[1]

[1]Unidad de Investigación, Hospital Galdakao-Usansolo, REDISSEC: Red de Investigación en Servicios Sanitarios y Enfermedades Crónicas,48960 Galdakao, Spain Spain.
[2]Departamento de Matemática Aplicada y Estadística e Investigación Operativa, Universidad del País Vasco (UPV/EHU), Bilbao, Spain.REDISSEC: Red de Investigación en Servicios Sanitarios y Enfermedades Crónicas, Spain.

Aims

Pre-post studies based on health related quality of life (HRQoL) variables are motivated to determine the potential predictors of the mean change of the outcome of interest. It is very common in such studies for data to be missing, which can bias the results. The appropriate statistical approach to analyze the whole sample, with non-ignorable missingness is a relevant issue that statisticians must address. Imputation techniques such as K-Nearest Neighbour (K-NN), Markov Chain Monte Carlo (MCMC) or Propensity score (PS) have been suggested as alternative to naive methods -Complete Case (CC), Available Case (AC)- to handle missing outcomes. The goal of the study was to compare the performance of various imputation techniques under different missingness mechanisms and rates.

Methods

Five analysis approaches - CC, AC, K-NN, MCMC and PS - combined with mixed models have been compared under different settings (rate: 10% and 30%; mechanisms: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR)). These strategies were applied to a pre-post study of 400 patients with chronic obstructive pulmonary disease (COPD). We analyzed the relationship of the changes in subjects HRQoL over one year with clinical and sociodemographic characteristics. A simulation study was performed (500 and 1000 runs), where the standardized bias of the regression coefficient of the interaction between the Time effect and the covariate was computed.

Results

In both 500 and 1000 simulation-runs, CC with mixed models showed the lowest standardized bias coefficients for MCAR and MAR scenarios. However, in MNAR setting, both approaches provided biased coefficients. PS was the worst imputation method.

Conclusions

MCMC has not additional benefit over CC when handling missing data for MCAR and MAR settings. In MNAR, all methods showed biased results.

References:

1. Altman DG, Bland JM (2007). Missing data. BMJ 334 (7590):424.
2. Barnard, J. and Meng, X. (1999) Applications of multiple imputation in medical studies: From AIDS to NHANES. Statistical Methods in Medical Research 8, 17-36.
3. Little, R.J.A. and Rubin, D.B. (2002) Statistical analysis with missing data. New York, Ed.

Un modelo de conglomerados con estructura de clases latentes para el análisis de mapas de enfermedades.

Karen C. Flórez -Lozano[1], Ana Corberán-Vallet[2], José D. Bermúdez [3]

[1]Departamento de Estadística e I.O, Universidad de Valencia, España y Departamento de Matemáticas y Estadística Universidad del Norte, Barranquilla- Colombia; [23]Departamento de Estadística e I.O, Universidad de Valencia, España.

Desde hace un poco más de dos décadas ha ido creciendo en el campo de la salud y específicamente en la Epidemiología un gran interés por estudiar la evolución de ciertas enfermedades que de algún modo representan una amenaza para la salud de las personas. Muchos investigadores han estado interesados en analizar la variación geográfica del suceso de salud en estudio y de cómo este se comporta a través de un período de tiempo, (Besag, York y Mollié ,1991; Lawson ,2009; entre otros).

En este trabajo presentamos un nuevo modelo que no requiere definir desde el inicio la dependencia o distancia entre vecinos, sino que expone una formulación donde variables de asignación de los riesgos permiten capturar diferentes estructuras de riesgo. Es un enfoque alternativo donde los riesgos relativos en áreas pequeñas son asignados a riesgos latentes. Nuestra propuesta aplica ideas de detección de conglomerados, modelos de mixturas y de modelos con estructura latente, (Knorr-Held y Rasser ,2000 ; Green y Richardson , 2002).

Proponemos un modelo bayesiano de conglomerados con estructura latente donde encontraremos variables observadas y no observadas. El vector de variables observadas son los conteos de la enfermedad en cada una de las áreas. Los parámetros del modelo son los riesgos latentes , el número de componentes, clases o conglomerados en los cuales se agruparán las áreas pequeñas según su riesgo y las probabilidades de pertenecer a cada componente. Consideramos también el número de conglomerados como otro parámetro desconocido.

Presentamos un estudio de simulación con el que pretendemos explorar el rendimiento del modelo propuesto. El principal objetivo de nuestro estudio es investigar la capacidad de recuperación del modelo verdadero y el impacto que se tiene dentro del mismo cuando componentes latentes no observables existen en el proceso y también cuando variamos la prevalencia de la enfermedad. También queremos estudiar el poder predictivo del mismo.

Referencias:

1. Besag J, York J, Mollié A.(1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*; 43: 1-59 (with discussion).
2. Lawson, A. (2009). *Hierarchical modeling in spatial epidemiology*. Chapman and Hall.

3. Knorr-Held, L. y Rasser, G. (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics*, 56(13-21):2045-2060.

Estudio de la mortalidad bovina registrada en granja mediante modelos INAR, potencial como componente de vigilancia sindrómica

Amanda Fernández[1] , Anna Alba[2], Pere Puig[1]

[1] Departament de matemàtiques, Universidat Autònoma de Barcelona, España
[2]Centre de Recerca en Sanitat Animal (CReSA), Fundació UAB-IRTA, Barcelona, España

La vigilancia de enfermedades en animales domésticos es fundamental para planificar e implementar medidas preventivas y de control, crear sistemas productivos eficientes y garantizar que los subproductos derivados sean saludables y seguros para el consumidor final.

El desarrollo de herramientas de minería de datos y análisis espacio-temporal permite obtener información sobre el estado de salud de una población a partir de datos inespecíficos como son los datos de la mortalidad registrados en granja. Estudios previos han demostrado que la mortalidad registrada en granjas de bovino presenta patrones característicos según la localización geográfica y el tipo de producción. Los análisis de series temporales a partir de estos datos y la predicción de valores esperados pueden servir como indicador de salud. No obstante estos patrones difieren según la subpoblación y a menudo es necesario estudiar estos perfiles a escala local en subpoblaciones muy pequeñas para poder adoptar medidas eficaces.

En el presente trabajo se analizan algunas de las series temporales registradas en subpoblaciones en las que las técnicas clásicas de series temporales podrían no ser adecuadas debido a que las observaciones son recuentos con valores bajos y con una proporción elevada de ceros. En este trabajo se presenta un método alternativo basado en los denominados modelos INAR, que son adecuados para modelar series temporales de recuento para patrones que siguen distribuciones de *Poisson*. El modelo INAR(k) se expresa mediante una ecuación en diferencias de tal manera que, dada la serie temporal $\{X_t, t \in \mathrm{Z}\}$, $\mathrm{X_t} = \mathrm{p_1}\circ\mathrm{X_{t-1}} + \mathrm{p_2}\circ\mathrm{X_{t-2}} + \cdots + \mathrm{p_k}\circ\mathrm{X_{t-k}} + \mathrm{W_t}$ donde $\circ$ es conocido como el operador *thinning* y $\mathrm{W_t} \sim \mathrm{Poisson}(\lambda)$. Esta notación significa que $\mathrm{p_j}\circ\mathrm{X_{t-j}} \sim \mathrm{Binomial}(\mathrm{X_{t-j}}, \mathrm{p\_j})$.

En este trabajo estudiaremos la modelización de tres series temporales recogidas en tres explotaciones y regiones diferentes, utilizando un modelo INAR(5) con $\lambda_t = e^{\beta_0+\beta_1 t+\beta_2 \cos\left(\frac{2\pi t}{52}\right)}$ para ajustar la tendencia y estacionalidad anual de la primera serie temporal, un modelo INAR(5) considerando $\lambda_t = e^{\beta_0+\beta_1 t}$ para la tendencia de la segunda serie y un modelo INAR(3) con $\lambda_t = e^{\beta_0+\beta_1 t+\beta_2 \cos\left(\frac{2\pi t}{52}\right)+\beta_3 \sin\left(\frac{2\pi t}{26}\right)}$ para modelar la tendencia y la estacionalidad anual y semestral de la tercera serie analizada.

Referencias:

1. *Alba A, Dorea FC, Arinero, L Sanchez J, Cord_on R, Puig P, Revie CW. Exploring the surveillance potential of mortality data: nine years of bovine fallen stock data collected in Catalonia (Spain) ICAHS, Proceedings ICAHS- 2nd International Conference on Animal Health Surveillance Cuba 7-9 May 2014, 129-130.*
2. Jung RC and Tremayne AR. Binomial thinning models for integer time series. Statistical Modelling (2006); 6: 81-96.
3. Moriña D, Puig P, R__os J, Vilella A and Trilla A. A statistical model for hospital admissions caused by seasonal diseases. Statistics in Medicine (2011); 30: 3125-3136.

Models for binary response and survival data

Ana Vázquez[1], Anna Espinal[1] , Olga Julià[2]

[1]Servei d'Estadística Aplicada, UAB; [2]Departament de probabilitat, lògica i estadística, UB

Usually time to event is measured on a continuous scale. However for several reasons time can be measured on a discrete scale. Discrete observed times often imply tied data. This characteristic requires specific methods of analyzing discrete survival data.

Models for binary response, with *logit* and *cloglog* link can be applied. But these models need an "extended dataset" build from the original dataset.

This database is defined from dummies variables at each uncensored time and the individual contributions at each of these times.

The advantages of using these models are that allows the possibility using all the methodologies and software for binary response models to deal with discrete survival data.

Estimations of the covariate effects from the proposed models are compared with the results obtained assuming a Proportional Hazards Model (Cox, 1972) taking into account tied data.

References:

1. Singer,J.D and Willett, J.B (2003). *Applied Longitudinal Data Analysis. Modeling change and event occurrence*. OXFORD University Express.

Spatial Disaggregation in Epidemiology

Diego Ayma Anza[1], María Durbán Reguera[1], Dae-Jin Lee[2]

[1]Universidad Carlos III de Madrid; [2]BCAM – Basque Center of Applied Mathematics

Abstract: Disease maps studies deal, in general, with aggregated data over geographical units like districts, city quarters or municipalities. These data are commonly displayed in choropleth maps that suffer from the biased visual perception generated by the differences in size and shape of geographical units. Also, epidemiologists are interested in relating these data with relevant risk factors, but they vary continuously or are measured in a fine grid. To overcome these drawbacks, we propose to use a generalization of the Penalized Composite Link Model (or P-CLM, Eilers, 2007) to the spatial case, into a mixed model framework (Lee and Durbán, 2009), to obtain a continuous surface that represents the underlying spatial trend at a finer scale. This model allows for different amount of smoothing for spatial coordinates at the fine scale, and parameter estimation is obtained via an iterative PQL procedure (Breslow and Clayton, 1993). The proposal methodology is applied to a dataset of the Community of Madrid, which come from a large European epidemiological study called MEDEA.

References:

1. Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. J. Am. Statis. Assoc. , 88:9-25.
2. Eilers, P.H.C. (2007). Ill-posed problems with counts, the composite link model and penalized likelihood. Statistical Modelling, 7:239-254.
3. Lee, D.-J., and Durbán, M. (2009) Smooth-CAR mixed models for count data. Computational Statistics & Data Analysis, 53:2958-2979.

Sesión Posters

Estudio del nomadismo y modelización estadística de la colonia de gaviota de Audouin de las Islas Columbretes

Blanca Sarzo[1], Carmen Armero[1], David Conesa[1] y Hèctor Perpiñán[1]

[1] Departamento de Estadística e Investigación Operativa (Universitat de València)

Resumen

La gaviota de Audouin (*Ichthyaetus audouinii*) es una especie endémica del Mediterráneo y emblemática para las Islas Columbretes. Desde el inicio de su colonización en los años 70, la colonia seguía un crecimiento creciente hasta alcanzar un máximo de 625 parejas reproductoras en el año 96. Este pico máximo se produjo a consecuencia de la entrada de un tejón en el año 94 en la colonia del Delta del Ebro, provocando la exportación de ejemplares a otras colonias cercanas. A partir de entonces, la colonia entró en un declive poblacional hasta la actual cifra de 10 parejas (dato relativo al año 2013).

Una de las características de la gaviota de Audouin es su carácter nomádico. La colonia, desde su establecimiento en el Archipiélago, ha cambiado de localización en las diferentes islas que componen el mismo. El objetivo de este estudio es analizar posibles patrones de ocupación de las islas por la colonia, e intentar entender cuáles son los factores relevantes en la dinámica poblacional de la especie. El estudio se ha realizado desde una perspectiva bayesiana a través de un modelo de regresión.

Los resultados obtenidos indican que sí parece existir cierto patrón de ocupación de las islas por parte de la especie, patrón que se ve truncado en varias ocasiones debido al establecimiento de la veda de arrastre (principal fuente de alimentación de la especie). En cuanto a la modelización estadística, los resultados sugieren que el número de parejas reproductoras de gaviota de Audouin no se ve afectado negativamente por la presencia en las islas de otra especie de gaviota competidora, la gaviota patiamarilla (*Larus michahellis*), como se había sugerido en anteriores estudios. A su vez, los resultados muestran la relevancia del carácter social de la especie para el futuro de la colonia, de manera que cuanto mayores sean los efectivos poblacionales de la especie el año anterior, mayor será el número de ejemplares que intentará la reproducción en las islas el año siguiente.

DESARROLLO DE MODELOS DE PREDICCIÓN DEL RIESGO DE ACCIDENTALIDAD BASADOS EN DATOS OBTENIDOS DE ENCUESTAS DE SEGURIDAD VIAL

Maider Mateo-Abad[1], Inmaculada Arostegui[1], Arantza Urkaregi[1]

[1]Departamento de Matemática Aplicada, Estadística e Investigación Operativa. UPV/EHU;

Financiación: Gobierno Vasco, Departamento de Interior (T-011/12) y Departamento de Educación, Política Lingüística y Cultura (IT620-13)

Abstract: up to 300 words.

Este trabajo fue realizado en conjunto entre la Universidad del País Vasco y la Dirección de Tráfico del Gobierno Vasco. El objetivo principal del proyecto es el desarrollo de un modelo predictivo del riesgo de accidentalidad en base a encuestas realizadas los años 2009 y 2010 por la Dirección de Tráfico de la CAPV para el estudio de conductores/as.

El objetivo principal consta de estudiar los factores predictivos para 1) riesgo de accidentalidad, donde la variable respuesta es dicótoma, Y=1, persona con algún accidente, Y=0, en caso contrario; 2) tasa de accidentalidad. Se utilizaron los modelos de regresión logística y regresión de Poisson para dar respuesta a los objetivos 1) y 2) respectivamente. Los resultados obtenidos con ambos modelos son muy similares, obteniendo como variables predictoras: sexo, edad junto a antigüedad del carnet, tiempo conducido, superar los límites de velocidad, ingerir alcohol y uso del móvil. Sin embargo, mediante la regresión de Poisson fue posible detectar algún factor más: conducción por excepción, tipo de conducción y uso del cinturón. Además, el modelo de Poisson, nos permite concluir que cuanto más joven y menos experiencia, la tasa de accidentalidad es mayor, mientras que en el modelo logístico, el riesgo de accidente aumenta según aumenta la edad y experiencia, debido a que la persona se expone al riesgo por más tiempo.

Adicionalmente se llevo a cabo un análisis exploratorio en el que se identificaron los individuos en base a la tipología de los conductores clasificándolos en 4 grupos: Personas Accidentadas por motivos no determinados (27.8%); Personas No Accidentadas, pero con mala conducta (24.4%); Personas Accidentadas con mala conducta (13.2%); Persona No Accidentadas (34.6%). La metodología utilizada fue un Análisis de Correspondencias Múltiple y Análisis de Conglomerados.

References:

1. Everitt, BS (2001). Cluster analysis. Arnold, London, 4th. ed.
2. Hilbe JM. (2011). Negative Binomial Regression. Cambridge, New York, 2nd ed.
3. Zuur A, Ieno E, Walker N, Saveliev A, Smith G (2009). Mixed Effects Models and Extensions in Ecology with R. Springer, New York

MODELIZACIÓN DE LA CALIDAD DE VIDA RELACIONADA CON LA SALUD MEDIANTE LA DISTRIBUCIÓN BETA-BINOMIAL: APLICACIÓN EN PACIENTES CON ENFERMEDADES CRÓNICAS

Najera J.[1], Arostegui I.[1]

[1]Universidad del País Vasco UPV/EHU. Departamento de Matemática Aplicada, Estadística e Investigación Operativa

La medición de la Calidad de Vida Relacionada con la Salud (CVRS) hace posible obtener información sobre la enfermedad y su impacto en la vida del paciente. El método clásico para determinar y evaluar de una forma estandarizada y objetiva el impacto de la enfermedad en la vida diaria y en la sensación de bienestar es la administración de cuestionarios.

La forma más común de analizar el efecto que tienen algunas características tanto del paciente como de la enfermedad en la CVRS es el modelo lineal general, mediante estimación por mínimos cuadrados. Dicha estimación exige la condición de distribución normal de la variable dependiente. Las puntuaciones de CVRS obtenidas mediante cuestionarios son, en general, medidas ordinales, con una distribución sesgada y acotada. Trabajos previos señalan lo inadecuado de las técnicas de modelización basadas en la distribución normal en este contexto (1). En este trabajo se propone la modelización de datos de CVRS mediante la distribución beta-binomial. El objetivo es ajustar la distribución beta-binomial a datos de CVRS y utilizar la regresión beta-binomial para modelizar estos datos.

Realizaremos una aplicación de la propuesta a datos de CVRS provenientes del Cuestionario de Salud SF-36 en pacientes con enfermedad pulmonar obstructiva crónica (EPOC). En primer lugar, ajustaremos los datos a la distribución normal, con lo cual mostraremos lo inapropiado de analizar este tipo de datos mediante modelos de regresión lineal. A continuación, mostraremos el ajuste de datos a la distribución beta-binomial y su modelización mediante regresión beta-binomial para cada una de las dimensiones del SF-36. Finalmente, mostraremos la interpretación clínica de los resultados. En resumen, en este trabajo mostraremos la superioridad del modelo de regresión beta-binomial frente al modelo lineal general clásico para modelizar datos de CVRS, tanto desde un punto de visto teórico como aplicado.

Financiación: 2012111008, IT620-13, MTM2013-40941-P, UFI11/52.

Referencias:

1. Arostegui I. Nuñez-Antón V. eta Quintana J.M. (2007). Analysis short form-36 (SF-36): The beta binomial distribution approach. *Statistics in Medicine*, **26**, 1318-1342.

Modelización de los residuos vegetales generados por las podas de palmeras como fuente de biomasa.

J.A. García-Gómez[1], X. Barber[1], R. Mora[l2]

[1]I.U.I Centro de Investigación Operativa.[2]Dpto. de Agroquímica y Medioambiente.

Universidad Miguel Hernández de Elche.

Corresponding author: codicealberto@gmail.com

Las especies palmáceas son un cultivo con una elevada importancia a nivel ambiental, paisajístico, económico y cultural, con una creciente presencia en la UE y especialmente en España. Además existen parajes y entornos protegidos de palmeras debido a su singularidad y riqueza en muchas áreas de Europa y también en España (sureste español y zonas insulares). La poda de estas especies representa una gran cantidad de biomasa por hectárea y año. Si consideramos una producción promedio de biomasa por poda de 5 kg anuales, obtenemos una producción de biomasa residual (mayoritariamente depositada en vertederos) cercana a 200.000 toneladas, sin considerar la biomasa residual adicional asociada a los especímenes afectados por la plaga del picudo rojo (Rhynchophorusferrugineus).

Se desea elaborar una imagen precisa de la situación de las especies de palmeras (Phoenix dactylifera, Phoenix canariensis, Washingtonia robusta) en Elche (Palmeral Patrimonio de la Humanidad) con el fin de estimar la producción de biomasa, incluida la generada por la infestación del Picudo Rojo (objetivo secundario que alcanzaríamos modelizando esta plaga). Con ello se conseguirá una caracterización de los residuos de palma, agrupados en partes vegetales (raíces, tallo, hojas y raquis) y así determinar sus propiedades físicas, químicas y biológicas y las posteriores opciones de valorización posibles.

Para este fin se van a utilizar modelos jerárquico Bayesiano espaciales y así poder predecir por un aparte la biomasa generada y el patrón de comportamiento de la infestación por Picudo Rojo en el Palmeral de Elche.

References:

1. S. Banerjee, B.P. Carlin, and A.E. Gerfand. Hierarchical Modeling and Analysis forSpatial Data. Chapman and Hall/CRC, 2014.
2. X. Barber. Modelos geoestadísticos para el estudio de indices bioblimaticos. PhDthesis, Universidad Miguel Hernández de Elche, 2009. URL http://goo.gl/s7xOMm.
3. Rubio, R., Pérez-Murcia, M.D., Agulló, E., Bustamante, M.A., Sánchez, C., Paredes, C., Moral, R.Recycling of Agro-food Wastes into Vineyards by Composting: Agronomic Validation in Field Conditions(2013) Communications in Soil Science and Plant Analysis, 44 (1-4), pp. 502-516.

Chufa negra, horchata blanca: búsqueda de soluciones para la mancha negra de la chufa

D. Alvares[1], C. Armero[1], A. Forte[1], L. Galipienso[2], A. Vicent[2] y L. Rubio[2]

[1]Universitat de València; [2]Instituto Valenciano de Investigaciones Agrarias

La chufa *Cyperus* sculentus es un tubérculo que se utiliza principalmente para la elaboración de horchata. Se cultiva exclusivamente en la comarca de la Horta Nord en Valencia y tiene una gran importancia socioeconómica. La mancha negra de la chufa es una enfermedad de origen desconocido: un porcentaje de los tubérculos cosechados sufre un ennegrecimiento de la piel que conlleva su depreciación comercial debido a que los tubérculos con mancha negra tienen que ser desechados.

El objetivo de este trabajo es analizar si la selección de tubérculos sin mancha para su utilización como simiente supone una mejora en la cosecha en cuanto a una mayor producción y tamaño de los tubérculos y una menor incidencia de la enfermedad.

Los datos analizados proceden de un experimento en invernadero con simientes de chufas sanas y enfermas, estas últimas clasificadas según síntomas leves y graves. Los análisis estadísticos se han realizado todos desde la metodología bayesiana. Los modelos lineales han sido la base para los análisis de la producción y tamaño de los tubérculos cosechados y los modelos multinomiales y logísticos multinomiales para la incidencia de la enfermedad.

Referencia:

Agresti, A. (2002), Categorical Data Analysis, Second Edition, New York: John Wiley & Sons.

Implementation of new methods of measurement and comparison of longevity in Europe

Patricia Carracedo [1], Ana María Debón[2]

[1,2]*Universitat Politècnica de València*

Mortality has decreased in all the countries of the European Union during the last century, presenting similar patterns within the change of mortality. Despite these similar trends, there are still considerable differences in the levels of mortality between eastern and western zone. The objective of this study is to present a method for detecting clusters (groups) of European countries with similar mortality. The method takes into account the geographical location of countries and, consequently, the neighbourhood relationships among them. Given the space-time structure of the mortality data, we have implemented panel data models in order to find interaction between the two components. The results confirm that there is space-time dependence, as previously exploratory analysis has confirmed.

References:

1. Debón, A., Montes, F., Martínez-Ruiz, F. 2011. Statistical methods to compare mortality for a group with non-divergent populations: an application to Spanish regions. European Actuarial Journal, 1, 291-308.
2. Rey, S. J. (2001). Spatial empirics for economic growth and convergence. Geographical Analysis, 33(3), 195-214.
3. Rey, S. J. (2014). Spatial Dynamics and Space-Time Data Analysis. In *Handbook of Regional Science* (pp. 1365-1383). Springer Berlin Heidelberg.

Rotavirus en niños de la Comunidad Valenciana: modelización espacial de las hospitalizaciones y de la cobertura vacunal

López Lacort M[1], Díez Domingo J[1] ,Pérez Vilar S[1],Martínez Beneito MA[1]

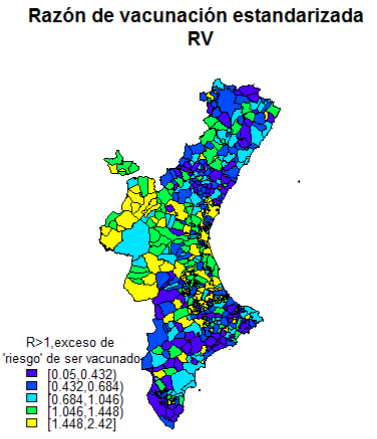
[1]FISABIO-Salud Pública, Valencia

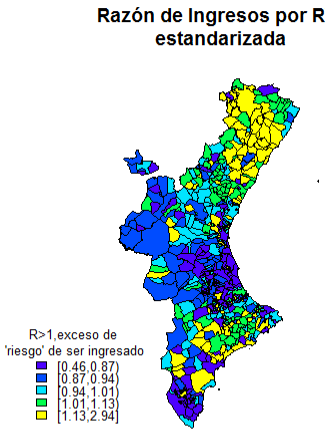
Rotavirus provoca diarreas en niños pequeños que frecuentemente se complican y requieren hospitalización. Para la Salud Pública Española la enfermedad producida por rotavirus no es un problema relevante, sin embargo los pediatras recomiendan la vacuna. Hay muy pocos estudios que valoren el impacto de la enfermedad y de la vacuna en nuestro medio, por lo que su análisis puede ayudar a la toma de decisiones de los programas de vacunación.

El objetivo de este estudio es describir la distribución geográfica de las gastroenteritis por rotavirus hospitalizadas y de la cobertura vacunal frente a rotavirus en niños de la Comunidad Valenciana, con la finalidad de valorar la relación entre ingresos y cobertura.

Para evaluar el comportamiento espacial de los ingresos por rotavirus y de la cobertura vacunal se han modelizado los estimadores del riesgo: la razón de ingresos estandarizada y la razón de vacunación estandarizada, suavizados mediante el modelo Besag-York-Mollié.

Nuestros resultados muestran una aparente relación inversa entre hospitalizaciones y cobertura vacunal (Figura 1 y Figura 2).





Un modelo bayesiano de respuesta multinomial y con valores perdidos para la caracterización de distintos tipos de neumonía en niños menores de 5 años.

López Lacort M[1], Martínez Beneito M.A[1], Díez Domingo J[1]

[1]FISABIO Salud Pública, Valencia.

La neumonía adquirida en la comunidad es una infección de los pulmones. Presenta una alta morbilidad en niños, especialmente en países en desarrollo. Su diagnóstico se basa en la clínica y la exploración radiológica del tórax, pero existe gran variabilidad en su interpretación, por lo tanto dar un diagnóstico consistente de neumonía es complicado.

El objetivo de este estudio es encontrar características que diferencien entre 3 tipos de neumonía: alveolar, no alveolar y clínica, con el fin de poder dar un diagnóstico más certero en la primera toma de contacto en la que no se dispone de una radiografía y en consecuencia ser más efectivos con la administración del tratamiento.

La metodología utilizada en el estudio consta del desarrollo de un modelo bayesiano de regresión logística multinomial que contempla como variables: neutrófilos en sangre, saturación de oxígeno, edad, estado vacunal frente neumococo y el efecto del hospital. Mediante técnicas de imputación múltiple en el mismo modelo hemos generado los datos faltantes de la base de datos, teniendo en cuenta que su ausencia no se debe en general al azar.

Las neumonías alveolares presentan niveles más altos de neutrófilos, la vacuna antineumocócica parece proteger de éstas confirmando su origen bacteriano. Las no alveolares ocurren en niños más pequeños , además por cada 10 % que desciende la saturación de O2 la probabilidad de tener una neumonía no alveolar se ve multiplicada por 2. Las edad y la saturación diferencian neumonías no alveolares de clínicas, en concreto las clínicas se dan en niños más mayores y el descenso de O2 es menor que en las no alveolares.

References:

1. Deshmukh Sachin, Manjrekar Pradip, Gopal R. A Brand choice model using multinomil logistic regression, bayesian inference and Marcov Chain Monte Carlo method Vol 1, Issue 1, 2010.

2. V Pando Fernández, R San Martín Fernández. Regresión logística multinomial. Cuad. Soc. Esp. Cien. For. 18(2004)

3. Carracedo-Martínez E, Figueiras A. Tratamiento estadístico de la fal-ta de respuestas en estudios epidemiológicos transversales.Salud Publica Mex 2006.

Modelización estadística del proteoma completocon células humanas en tejido normal y tumoral.

Raquel Gavidia Josa, Carmen Armero, Luz Valero y Manuel Sánchez

Objetivo.
Las investigaciones proteómicas tratan de identificar y cuantificar las variaciones de abundancia observadas en una proteína o grupos de proteínas entre individuos sanos y enfermos con el fin de descubrir biomarcadores de pronóstico o diagnóstico.
El objetivo de este trabajo es encontrar qué proteínas diferencian el tejido sano del tejido tumoral.

Descripción de los datos
Los datos provienen de un estudio realizado en mujeres enfermas de cáncer de mama. La técnica utilizada para identificar y cuantificar ha sido el marcaje i-TRAQ. Permite marcar como máximo 8 muestras por experimento. Tenemos un total de 2040 proteínas y 8 muestras que consideraremos independientes.

Métodos usados para el análisis
Hemos seguido tres líneas de investigación:

1. 1. Contrastes múltiples corregidos por FDR, controla la proporción de falsos positivos entre las hipótesis rechazadas. Tenemos el problema de trabajar con muchas más variables que muestras u observaciones.

Alternativas:

1. 2. Modelo regresión con restricción Lasso: somos capaces de decir qué proteínas nos diferencian mejor los tejidos sanos de los enfermos, puesto que penaliza los coeficientes de las variables que no están asociadas con la variable respuesta a cero.
2. 3. Método de reducción de dimensión PLS-DA hace una buena clasificación de las muestras de tejidos, en donde todas las variables están proyectadas en los componentes que crea el modelo.

Conclusiones
Lasso identifica cuatro proteínas siendo tres de ellas iguales a las obtenidas mediante contrastes múltiples. La proteína PTBP1 aparece en un artículo publicado en un estudio relativo al cáncer de mama.

References:

1. 1. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New York: Springer, 2009 (Vol. 2, No. 1).

2. 2. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, 1995

Distribución climática del citrus black spot causado por *Phyllosticta citricarpa*. Un análisis descriptivo de la expansión de la enfermedad en Sudáfrica.

Joaquín Martínez-Minaya[1], David Conesa[2], Antonio López-Quílez[2], Antonio Vicent[1]

[1] Centro *de Protección Vegetal y Biotecnología, Instituto Valenciano de Investigaciones Agrarias (IVIA), Moncada, 46113 Valencia, Spain.*

[2] *Departament d'Estadística i Investigació Operativa, Universitat de Valéncia, C/ Dr. Moliner 50, Burjassot, 46100 Valencia, Spain.*

Abstract

En el siguiente póster se propone una aproximación metodológica para la modelización de patrones de ocurrencia de una enfermedad (Citrus black spot). Citrus black spot (CBS) es una enfermedad causada por el hongo *Phyllosticta citricarpa* (McAlpine) Van der Aa (syn. *Guignardia citricarpa* Kiely). CBS fue detectada por primera vez en Australia, y, actualmente está distribuida en los países citrícolas de África, Sudamérica, EE.UU., y China. España y el resto de países de la cuenca del Mediterráneo están exentos de la enfermedad y *P. citricarpa* está considerado como organismo de cuarentena en todo el territorio de la Unión Europea (UE). Así, lo que se pretende con este estudio es determinar la importancia de los factores climáticos y espaciales en la distribución geográfica del 'citrus black spot' en Sudáfrica, con el fin de estimar el riesgo de establecimiento de la enfermedad en las zonas citrícolas de la Comunitat Valenciana. Para ello, las principales fuentes de datos que se emplearán son la base de datos WorldClim y la cartografía del CBS en Sudáfrica.

Dada la variedad climática encontrada en Sudáfrica, se han elaborado algoritmos para obtener la clasificación climática de Koppen y la clasificación climática de Aschmann de las zonas de dicho país. Además, se ha comprobado la presencia de autocorrelación espacial en los datos mediante el estadístico de Moran y de Geary.

Por último, se está realizando un modelo jerárquico bayesiano espacial empleado para modelizar la presencia/ausencia de dicha enfermedad, donde se están usando factores climáticos y geográficos de cada localización donde está presente el hongo. La inferencia bayesiana en los parámetros se está realizando considerándolo como un modelo latente Gausiano, el cual permite el uso de INLA (Integrated Nested Laplace Approximation Software).

References:

1. Hijmans, R. J. (2014). raster: raster: Geographic data analysis and modeling. R package version 2.2-31. *http://CRAN.R-project.org/package=raster*.
2. Makowski, D., Vicent, A., Pautasso, M., Stancanelli, G., & Rafoss, T. (2014). Comparison of statistical models in a meta-analysis of fungicide treatments for the control of citrus black spot caused by *Phyllosticta citricarpa*. *European Journal of Plant Pathology, 139*, 79-94.
3. Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Koppen-Geiger climate classification. *Hydrology and Earth System Sciences, 11*, 1633-1644.

Using beta regressions to study discard rates in fisheries

Marcial Marín[1], Iosu Paradinas[1], Jose Mª Bellido[2], David V. Conesa[1]

[1]Universitat de València

[2]Instituto Español de Oceanografía

Discards (removing part of the catches) are an inherent part of fishery activity. Their perception as a negative factor for efficient use of resources has led to changes in European legislation for this sector. Appropriate studies to properly assess the impact of fisheries on stocks are needed, and the study of discards must be part of them. Works that model the discarded amounts are based on standardized measures as LPUE (Discards Per Unit Effort) and discards rates, mainly using linear and logistic regressions. Because of characteristics of discard rate (continuous variable that takes values between 0 and 1) we consider that a beta distribution may be appropriate to model this variable. In this poster we use beta regression models for modeling the rate of discards in fisheries.

Spatio-temporal classification in point patterns under the presence of clutter

Delfín Carot[1], *Jorge Mateu*[1]

[1]*Department of Mathematics, University Jaume I, Castellón*

Abstract.

We consider the problem of detection of features in the presence of clutter for spatio-temporal point patterns. This problem was previously treated but only in the spatial context. In particular, Byers et al. (1998) used k-th nearest neighbour distances to classify points between clutter and features. They proposed a mixture of distributions whose parameters were estimated using an EM algorithm.
This paper extends this methodology to the spatio-temporal context by considering the properties of the spatio-temporal k-th nearest neighbour distances. We make use of several spatio-temporal n-dimensional distances (n-1 spatial dimensions, and 1 temporal dimension), that are mixtures of defined distances for the p-norm. We show close forms for the probability distributions of such k-th nearest neighbour distances. We also present an intensive simulation study that covers a wide range of practical scenarios; and an application to earthquakes in Andalucía.

References:

1. Cressie, N. (1993). *Statistics for spatial data*, Revised Edition. Wiley, New York.
2. Byers, S.D. & Raftery, A.E. (1998). Nearest Neighbor Clutter Removal for Estimating Features in Spatial Point Processes. *Journal of the American Statistical Association*, **93**, pp. 557-584.
3. Peng, R.D., Schoenberg, F.P. & Woods, J.A. (2005). *A Space-Time Conditional Intensity Model for Evaluating a Wildfire Hazard Index*. Journal of the American Statistical Association, **100**, 26-35.

La huella ecológica corporativa de los muncipios españoles.

Alejandro Martínez Gascón[1]

[1]*Author 1 Estudiante de la Universidad de Valencia*

El trabajo de fin de máster que expongo trata de calcular la huella ecológica corporativa bruta de los presupuestos consolidados y liquidados de los municipios españoles en 2010 medida en hectáreas globales por habitante. El siguiente paso es hacer inferencia sobre cual es la media nacional, autonómica y provincial. A través de la exportación de la idea usada en epidemiología de estandarizar el comportamiento local al nacional y, además, de una decisión para la cual se tiene en cuenta la técnica de simular se hace posible aprovechar la información que publica el Ministerio de Hacienda y Administraciones Públicas en Internet. Otra llave clave la podemos encontrar en los trabajos de Juan Luis Doménech Quesada, Adolfo Carballo Penela y Cabonfeel, sin olvidar a los autores Mathis Wackernagel y William Rees. Gracias a la potencia de computación de R se puede hacer el cálculo en poco tiempo. A fecha de hoy ya se dispone de los resultados de más de 7000 municipios donde están incluidos la grandes capitales. Queda trabajo por hacer, pero ya se empieza a ver la luz... Debo decir también que trabajos así aún no se han realizados todavía.

References:

1. Rees, W. Wackernagel, M. (2001) “Nuestra huella ecológica: reduciendo el impacto humano sobre la Tierra.” LOM. Santiago de Chile.
2. Doménech Quesada, J.L. (2009) “Huella ecológica y desarrollo sostenible.” Aenor. Madrid.
3. Carballo Penela, Adolfo. (2010) "Ecoetiquetado de bienes y servicios para un desarrollo sostenible" Aenor. Madrid.