

QÜESTIÓ, vol. 26, 1-2, pp. 109-122, 2002

MS

Análisis factorial múltiple como técnica de estudio de la estabilidad de los resultados de un análisis de componentes principales***Multiple Factor Analysis as a statistical technique to study the stability of the results obtained by Principal Component Analysis***

E. Abascal* y M. I. Landaluce**

* Departamento de Estadística e Investigación Operativa. Universidad Pública de Navarra.

** Departamento de Economía Aplicada. Universidad de Burgos.

Una característica de los métodos factoriales es que siempre producen resultados y éstos no son una simple descripción, sino que ponen de manifiesto la estructura existente entre los datos, de ahí la necesidad de estudiar la validez de los resultados. Es preciso analizar la naturaleza de esta estructura y estudiar la estabilidad de los resultados. Consideramos que el mejor criterio es el análisis de la estabilidad de los mapas obtenidos en el análisis factorial. El Análisis Factorial Múltiple (AFM), desarrollado por B. Escofier and J. Pagès (1992), permite el análisis simultáneo de tablas en las que un mismo conjunto de individuos se describe a través de varios grupos de variables. En este trabajo se muestra la eficacia de este método para el análisis de la estabilidad de los resultados obtenidos mediante Análisis de Componentes Principales.

Palabras clave: Análisis factorial múltiple, análisis de componentes principales y estabilidad.



QÜESTIÓ, vol. 23, 3, pp. 561-571, 1999

GE/BI

Generalization of the kappa coefficient for ordinal categorical data, multiple observers and incomplete designs

V. Abraira* and A. Pérez de Vargas**

* Unit of Clinical Biostatistics. Hospital 'Ramón y Cajal'. Madrid.

** Department of Biomathematics. University Complutense of Madrid.

This paper presents a generalization of the kappa coefficient for multiple observers and incomplete designs. This generalization involves ordinal categorical data and includes weights which permit pondering the severity of disagreement. A generalization for incomplete designs of the kappa coefficient based on explicit definitions of agreement is also proposed. Both generalizations are illustrated with data from a medical diagnosis pilot study.

Keywords: Agreement, kappa, incomplete designs.



QÜESTIIÓ, vol. 25, 1, pp. 121-156, 2001

BI

Mesa Redonda: la Estadística en la Investigación Médica

Round Table: Statistics in Medical Research

V. Abraira, C. Cadarso, G. Gómez, A. Martín y S. Pita

Este artículo es una transcripción de las conferencias dictadas en la mesa organizada en el 4º Congreso Galego de Estatística e Investigación de Operacions que tuvo lugar en Santiago de Compostela en Noviembre de 1999. Los autores discuten sobre el posible uso o abuso de la estadística en artículos científicos, sobre lo que se necesitaría para alcanzar la interdisciplinariedad y lo que se entiende por éxito profesional. Se define la disciplina (Bioestadística) y se identifica a sus profesionales (Bioestadísticos). Se discute sobre el papel de un bioestadístico en un equipo de investigación médico y se repasan las dificultades que tienen los médicos para realizar estudios clínico-epidemiológicos.

Palabras clave: Colaboración bioestadística, controversias bioestadísticas, multidisciplinariedad, reconocimiento profesional.



QÜESTIIÓ, vol. 24, 1, pp. 27-51, 2000

TS

Spatial structure analysis using planar indices

J. M. Albert*, J. Mateu** and J. C. Pernías**

* Department of Economics. Universitat Jaume I. Castelló.

** Department of Mathematics. Universitat Jaume I. Castelló.

Spatial planar indices have become a useful tool to analyze patterns of points. Despite that, no simulation study has been reported in literature in order to analyze the behaviour of these quantities under different pattern structures. We present here an extensive Monte Carlo simulation study focused on two important indices: the Index of Dispersion and the Index of Cluster Size, usually used to detect lack of homogeneity in a spatial point model. Finally, an application is also presented.

Keywords: Index of cluster size, index of dispersion, point processes, spatial structure.



QÜESTIIÓ, vol. 25, 1, pp. 71-90, 2001

GT

El índice de poder de Banzhaf en la Unión Europea ampliada

The Banzhaf power index in the EU after the enlargement

E. Algaba, J. M. Bilbao, J. R. Fernández y J. J. López

Matemática Aplicada II. Escuela Superior de Ingenieros. Universidad de Sevilla.

En este trabajo se definen algoritmos, basados en funciones generatrices, para calcular el índice de poder de Banzhaf en juegos simples de votación ponderada y en juegos de doble y triple mayoría. La utilización de funciones generatrices permite un cálculo exacto del índice de Banzhaf con una reducción sensible de la complejidad temporal. Además se calculan los índices de Banzhaf para las reglas de decisión, aprobadas en la cumbre de Niza, que se utilizarán en la Unión Europea ampliada a 27 países. Finalmente, se demuestra que los sistemas de triple mayoría adoptados son equivalentes en la práctica a juegos de mayoría simple o doble, porque la cuota de población exigida para aprobar una decisión no cambia el índice de Banzhaf de los países de la Unión Europea ampliada.

Palabras clave: Funciones generatrices, índice de Banzhaf, juegos de votación.

○○○

SORT, vol. 31, 2, pp. 109-150, 2007

AO

On modelling planning under uncertainty in manufacturing

A. Alonso-Ayuso*, L. F. Escudero* and M. T. Ortuño**

* Dpto. de Estadística e Investigación Operativa, Universidad Rey Juan Carlos.

** Dpto. de Estadística e Investigación Operativa, Universidad Complutense de Madrid.

We present a modelling framework for two-stage and multi-stage mixed 0-1 problems under uncertainty for strategic Supply Chain Management, tactical production planning and operations assignment and scheduling. A scenario tree based scheme is used to represent the uncertainty. We present the Deterministic Equivalent Model of the stochastic mixed 0-1 programs with complete recourse that we study. The constraints are modelled by compact and splitting variable representations via scenarios.

Keywords: Supply chain, BoM, strategic planning, scheduling, uncertainty, stochastic programming, Branch-and-Fix Coordination.

○○○

QÜESTIÓ, vol. 25, 3, pp. 479-498, 2001

MA

La minería de datos, entre la estadística y la inteligencia artificial

Data mining, between statistics and artificial intelligence

T. Aluja

Departamento de Estadística e Investigación Operativa. Universitat Politècnica de Catalunya.

En la pasada década hemos asistido a la irrupción de un nuevo concepto en el mundo empresarial: el "data mining" (minería de datos). Algunas empresas han implementado unidades de minería de datos estrechamente vinculados a la dirección de la empresa y en los foros empresariales las sesiones dedicadas a la minería de datos han sido las protagonistas. La minería de datos se presenta como una disciplina nueva, ligada a la Inteligencia Artificial y diferenciada de la Estadística. Por otro lado, en el mundo estadístico más académico, la

minería de datos ha sido considerada en su inicio como una moda más, aparecida después de los sistemas expertos, conocida desde hacía tiempo bajo el nombre de “data fishing”. ¿Es esto realmente así? En este artículo abordaremos las raíces estadísticas de la minería de datos, los problemas que trata, haremos una panorámica sobre el alcance actual de la minería de datos, presentaremos un ejemplo de su aplicación en el mundo de la audiencia de televisión y, por último, daremos una visión de futuro.

Palabras clave: Data mining, análisis de datos, modelización, inteligencia artificial, KDD, redes neuronales, árboles de decisión.

○○○

QÜESTIIÓ, vol. 24, 2, pp. 267-291, 2000

TS

Selección de la ventana en suavización tipo núcleo de la parte no paramétrica de un modelo parcialmente lineal con errores autorregresivos

Bandwidth selection in kernel smoothing of the nonparametric part in a partial linear model with autoregressive errors

G. Aneiros

Departamento de Matemáticas. Facultad de Informática. Universidad de La Coruña.

Supongamos que $y_i = \zeta_i^T \beta + m(t_i) + \varepsilon_i$, $i = 1, \dots, n$, donde el vector $(p \times 1) \beta$ y la función $m(\cdot)$ son desconocidos, y los errores ε_i provienen de un proceso autorregresivo de orden uno (AR(1)) estacionario. Discutimos aquí el problema de la selección del parámetro ventana de un estimador tipo núcleo de la función $m(\cdot)$ basado en un estimador Generalizado de Mínimos Cuadrados de β . Obtenemos la expresión asintótica de una ventana óptima y proponemos un método para estimarla, de modo que dé lugar a un estimador óptimo de $m(\cdot)$. Los resultados obtenidos generalizan aquellos obtenidos por Quintela (1994b) en regresión no paramétrica.

Palabras clave: Modelos parcialmente lineales, suavización núcleo, selección de la ventana, series de tiempo.

○○○

QÜESTIIÖ, vol. 22, 2, pp. 253-289, 1998

RE

Reflexiones sobre la estrategia de medida de los cambios en probabilidad en modelos de elección binarios

Reflections on the probability change measurement strategy in binary choice models

M. T. Aparicio e I.Villanúa

Departamento de Análisis Económico. Facultad de Ciencias Económicas y Empresariales. Universidad de Zaragoza.

Este trabajo se centra en la evaluación de la medida que, en el marco de los modelos de elección binarios o dicotómicos, se utiliza para reflejar el cambio en la probabilidad ante la variación de una de las variables explicativas. La opción de cuantificación más común ha consistido en utilizar el vector de valores medios de las variables explicativas, lo que podemos entender como poner el énfasis en el comportamiento de “un individuo medio”. Frente a esta práctica habitual, efectuamos una propuesta de evaluación alternativa que atiende fundamentalmente al “comportamiento medio de la muestra”. Se diseña, asimismo, un experimento de Monte Carlo que permite probar que la propuesta desarrollada en este trabajo es más robusta que la tradicionalmente planteada en la literatura.

Palabras clave: Elección discreta, variación de probabilidad, robustez.



QÜESTIÓ, vol. 22, 3, pp. 529-538, 1998

BI

The usefulness of discrimination based on distances on human evolution

C. Arenas* and D. Turbón**

* Departament d'Estadística. Universitat de Barcelona. Facultat de Biologia. Universitat de Barcelona.

** Departament de Biologia Animal (Antropologia). Facultat de Biologia. Universitat de Barcelona.

The reconstruction of human history from the fossil record often runs up against incomplete or differential preservation of specimens. In anthropological studies a large number of variables are usually taken and missing values can be a problem. Here we analyze three population samples of extinct aborigines from Tierra del Fuego. The first sample, with sex and ethnic group known, is used to compare the step-wise discriminant analysis and the discriminant analysis based on distances. With the second sample a first approach to the assignation of poorly documented specimens in relation to sex or ethnic group is presented here by comparing the results from the two discriminant methods. A third sample of skulls with ethnic group and sex unknown is used to illustrate the advantages of distance based discriminant analysis to solve the problem of allocating individuals when some values are missing.

Keywords: Distance discrimination, step-wise discrimination, Missing values, Tierra del Fuego aborigines.



SORT, 29, 1, pp. 119-138, 2005

SM

The M/G/1 retrial queue: An information theoretic approach

J. R. Artalejo and M. J. López Herrero
Complutense University of Madrid

In this paper, we give a survey of the use of information theoretic techniques for the estimation of the main performance characteristics of the M/G/1 retrial queue. We focus on the limiting distribution of the system state, the length of a busy period and the waiting time. Numerical examples are given to illustrate the accuracy of the maximum entropy estimations when they are compared versus the classical solutions.

Keywords: Principle of maximum entropy, M/G/1 retrial queue, limiting distribution, busy period, waiting time.

○○○

QÜESTIÓ, vol. 25, 1, pp. 3-17, 2001

ST

Contribuciones al muestreo sucesivo: estimador producto multivariante

Successive sampling using a multivariate product estimate

E. M. Artés Rodríguez

Departamento de Estadística y Matemática Aplicada. Universidad de Almería.

Se considera el problema de estimar la media de una población finita para la ocasión actual, basándonos en las muestras seleccionadas en dos ocasiones. Se construye un estimador producto multivariante de doble muestreo para la parte solapada de la muestra, para el caso en que dos variables auxiliares se encuentran correlacionadas de forma negativa con la variable objeto de estudio. Se obtienen las expresiones para el estimador óptimo y su varianza. Se calcula la curva que nos proporciona la ganancia en eficiencia del estimador combinado sobre el estimador directo que no utiliza la información obtenida en la primera ocasión. Se obtienen las condiciones bajo las cuales nuestro estimador mejora en precisión al estimador combinado de producto univariante. Finalmente, se incluye un estudio empírico para analizar el buen funcionamiento del método propuesto.

Palabras clave: Muestreo en ocasiones sucesivas, estimador producto bivariante, ganancia en eficiencia, fracción de solapamiento.

○○○

QÜESTIÓ, vol. 26, 1-2, pp. 197-211, 2002

OS

Smoothing the Catalan tourism micro-data time series

M. Artís*, J. L. Carrión*, À. Costa** and J. Suriñach*

* Anàlisi Quantitativa Regional (AQR) Research Group Departament d'Econometria, Estadística i Economia Espanyola Universitat de Barcelona.

** Institut d'Estadística de Catalunya.

In this paper we propose a method for smoothing the Catalan tourism time series between 1997 and 2000. These time series, built upon a micro database drawn from a survey conducted by the Statistical Institute of Catalonia, are somewhat volatile due, it would seem, to the incomplete nature of the information. The application of a smoothing procedure based

on the combination of classical techniques and weighted moving averages allows us to overcome the problems caused by this lack of information and to obtain time series that evolve smoothly over time.

Keywords: Smoothing micro-data, tourism time series.



QÜESTIÓ, vol. 26, 1-2, pp. 247-258, 2002

BI

Maintenance policy under unrevealed failures

F. G. Badía, M. D. Berrade and C. A. Campos
Departamento de Métodos Estadísticos. Centro Politécnico Superior de Ingenieros. Universidad de Zaragoza.

The unrevealed failures of a system are detected only by inspection. In this work, an inspection policy along with a maintenance procedure for multiunit systems with dependent times to failure is presented. The existence of an optimum policy is also discussed.

Keywords: Maintenance, reliability, replacement, optimum policy.



QÜESTIÓ, vol. 24, 3, pp. 531-552, 2000

OS

Avaluació dels mercats dels serveis d'informació electrònica

Electronic information services markets assessment

T. Baiget
Institut d'Estadística de Catalunya

Es presenten les característiques i els principals resultats del "MSStudy II", un projecte per avaluar la situació dels mercats dels serveis d'informació electrònica a l'Àrea econòmica europea (AEE), [Study for assessing the situation of the markets for electronic information services in the European economic area (EEA)] que, impulsat per la Direcció General XIII de la Comissió Europea, ha dut a terme a Espanya l'Institut d'Estadística de Catalunya (Idescat), amb la col·laboració de l'Asociación Multisectorial de la Información (Asedie). El MSStudy II, que durant 1998 i 1999 s'ha realitzat simultàniament en 16 països europeus, és una activitat estadística experimental recollida, però, en el Pla Estadístic de Catalunya 1997-2000 i ha tractat d'obtenir dades desglossades per tipus d'informació, tipus de suport o mitjà pel qual es distribueix la informació, tipus d'usuari, etc. Els resultats de cada país han estat integrats l'estiu del 2000 per un equip de professionals alemanys –els mateixos que varen planificar i dirigir aquest exercici per encàrrec de la CE.

Paraules clau: Mercat, informació electrònica, programes europeus, societat de la informació, tecnologies de la informació, continguts informatius.



QÜESTIÓ, vol. 23, 2, pp. 297-320, 1999

OS

Enlace de encuestas: una propuesta metodológica y aplicación a la encuesta de presupuestos de tiempo

Linking surveys: a method and application to the Survey of Time Uses

M. J. Bárcena y F. Tusell

Departamento de Estadística y Econometría. Facultad de CC.EE. y Empresariales. Bilbao.

Tratamos el problema de completar dos ficheros con registros conteniendo un subconjunto común de variables. La técnica investigada utiliza árboles de regresión y/o clasificación. Se propone y estudia una extensión para variables respuesta multivariantes, ilustrando su empleo sobre la Encuesta de Presupuestos de Tiempo (EPT-93).

Palabras clave: Enlace de ficheros, inserción de encuestas, imputación, árboles de regresión y clasificación.

○○○

SORT, vol. 30, 2, pp. 193-204, 2006

SI

Statistical models to study subtoxic concentrations for some standard mutagens in three colon cancer cell lines

X. Bardina*, L. Fernández**, E. Piñeiro**, J. Surrallés** and A. Velázquez**

* Departament de Matemàtiques, Universitat Autònoma de Barcelona.

** Departament de Genètica, Universitat Autònoma de Barcelona.

The aim of this work is to propose models to study the toxic effect of different concentrations of some standard mutagens in different colon cancer cell lines. We find estimates and, by means of an inverse regression problem, confidence intervals for the subtoxic concentration, that is the concentration that reduces by thirty percent the number of colonies obtained in the absence of mutagen.

Keywords: Inverse regression problem, subtoxic concentration, confidence interval.

○○○

QÜESTIÓ, vol. 22, 2, pp. 365-378, 1998

BI

Analysis of survey data investigating the malarial endemicity of a mixed tribal population of Bihar, India

T. K. Basu*, S. Ganguly*, S. K. Sarkar* and R. Arnab**

* Biometry Research Unit. Indian Statistical Institute. India.

** Department of Statistics. University of Durban, South Africa.

A section of the mixed tribal population of the Singhbhum district, Bihar, India is declared malaria epidemic zone. The tribal population of several generation is known to be suffering from malaria. A survey based on crosssectional data analysis, was conducted on the mixed tribal population for one month. The purpose of this study was to investigate the health status using collected blood samples. The main focus is on the comparative roles of the "defense mechanism" and "vitality factor" of the human system in context to the malarial infection. By gradual elimination of the blood parameters by statistical analyses, the "vitality" parameters probing malarial endemicity are assessed with a view to predicting the epidemic. The main findings from this survey are (i) of the selected twenty two parameters of blood, albumin, total cholesterol, total protein, b-Globulin, g-Globulin, Immuno globulin G seem to have some predictive capacity with respect to the malarial endemicity of the tribal people and (ii) Categorical variables like blood groups and sex are comparatively less important for prediction of malaria.

Keywords: Vitality factors, defense mechanism, two-way Anova, discriminant function.



QÜESTIÓ, vol. 23, 1, pp. 85-109, 1999

NO

Minimización global de un polinomio en la recta real

Globally minimizing a polynomial on the real line

C. Beltrán Royo

Dept. d'Estadística i Investigació Operativa. Secció d'Informàtica. Universitat Politècnica de Catalunya.

En este artículo presentamos y probamos numéricamente un nuevo algoritmo para la minimización global de un polinomio de grado par. El algoritmo está basado en la simple idea de trasladar verticalmente el grafo del polinomio hasta que el eje OX sea tangente al grafo del polinomio trasladado. En esta privilegiada posición, cualquier raíz real del polinomio trasladado es un mínimo global del polinomio original.

Palabras clave: Optimización global, máximo común divisor de polinomios, algoritmo de Euclides, secuencia de Sturm, división sintética de dos polinomios.



QÜESTIÓ, vol. 24, 2, pp. 251-266, 2000

RE

Comparación de curvas de supervivencia gamma estocásticamente ordenadas

Comparison of stochastically ordered gamma survival curves

J. D. Bermúdez* y E. Beamonte**

* Departamento de Estadística e Investigación Operativa. Universitat de València.

** Departamento de Economía Aplicada. Universitat de València.

En este trabajo se propone un análisis de supervivencia basado en un modelo Gamma. Se obtienen las condiciones teóricas bajo las cuales dos funciones de supervivencia Gamma están estocásticamente ordenadas. Estos resultados se utilizan para proponer un método sencillo que permite comparar dos poblaciones cuando, a priori, se conoce que sus curvas de supervivencia están estocásticamente ordenadas. Los resultados se ejemplifican con el análisis de un banco de datos reales sobre tiempos de desempleo.

Palabras clave: Supervivencia bayesiana, información inicial, métodos de Monte Carlo, predicción.

○○○

SORT, vol. 31, 1, pp. 3-44, 2007

BS

Objective Bayesian point and region estimation in location-scale models

J. M. Bernardo
Universitat de València.

Point and region estimation may both be described as specific decision problems. In point estimation, the action space is the set of possible values of the quantity on interest; in region estimation, the action space is the set of its possible credible regions. Foundations dictate that the solution to these decision problems must depend on both the utility function and the prior distribution. Estimators intended for general use should surely be invariant under one-to-one transformations, and this requires the use of an invariant loss function; moreover, an objective solution requires the use of a prior which does not introduce subjective elements. The combined use of an invariant information-theory based loss function, the intrinsic discrepancy, and an objective prior, the reference prior, produces a general solution to both point and region estimation problems. In this paper, estimation of the two parameters of univariate location-scale models is considered in detail from this point of view, with special attention to the normal model. The solutions found are compared with a range of conventional solutions.

Keywords: Confidence Intervals, Credible Regions, Decision Theory, Intrinsic Discrepancy, Intrinsic Loss, Location-Scale Models, Noninformative Prior, Reference Analysis, Region Estimation, Point Estimation.

○○○

QÜESTIÓ, vol. 24, 2, pp. 317-331, 2000

NF

Graphs, topologies and simple games

J. M. Bilbao
Matemática Aplicada II, Escuela Superior de Ingenieros. Universidad de Sevilla.

We study the existence of connected coalitions in a simple game restricted by a partial order. First, we define a topology compatible with the partial order in the set of players. Second, we prove some properties of the covering and comparability graphs of a finite poset. Finally,

we analyze the core and obtain sufficient conditions for the existence of winning coalitions such that contains dominant players in simple games restricted by the connected subspaces of a finite topological space.

Keywords: Finite poset, covering and comparability graphs, simple games.



QÜESTIÓ, vol. 23, 3, pp. 525-541, 1999

OS

La articulación entre lo cuantitativo y lo cualitativo: de las grandes encuestas a la recogida de datos intensiva

The articulation between the quantitative and the qualitative: from the great surveys to the intensive data collection

V. Borràs, P. López y C. Lozares

Departament de Sociologia, Universitat Autònoma de Barcelona.

Son casi innumerables las reflexiones que se han hecho en el campo de la metodología de las ciencias sociales, sobre la dicotomía, real o inexistente, entre las perspectivas de análisis cuantitativo y cualitativo, mientras que han sido menos abundantes los trabajos teóricos, aunque van siendo más frecuentes los empíricos, que han tratado de compatibilizar y/o complementar ambas perspectivas. En muchos casos los trabajos cualitativos cubren solamente los primeros pasos de la investigación social, como fases previas o exploratorias de la misma, para prolongarla después con una metodología más cuantitativa y extensiva como la inherente a las encuestas, al análisis de datos, multivariados, causales u otros. Este artículo presenta un ejemplo concreto del planteamiento inverso. Se parte de una gran encuesta con el objetivo de estructurar la realidad objeto de estudio, para posteriormente aplicar técnicas cualitativas, concretamente grupos de discusión.

Palabras clave: Análisis de datos, análisis de correspondencias, análisis de clasificación, metodología, consumo, técnicas cuantitativas y cualitativas.



SORT, vol. 28, 1, pp. 37-54, 2004

TS

Local superefficiency of data-driven projection density estimators in continuous time

D. Bosq and D. Blanke

Université Pierre et Marie Curie, Paris.

We construct a data-driven projection density estimator for continuous time processes. This estimator reaches superoptimal rates over a class F_0 of densities that is dense in the family of all possible densities, and a “reasonable” rate elsewhere. The class F_0 may be chosen previously by the Statistician. Results apply to IRd-valued processes and to N-valued

processes. In the particular case where square-integrable local time does exist, it is shown that our estimator is strictly better than the local time estimator over F_0 .

Keywords: Density estimation, data-driven, superefficiency, continuous time processes.

○○○

QÜESTIÓ, vol. 26, 1-2, pp. 3-14, 2002

PD

A note on the application of integrals involving cyclic products of kernels

V. Buldygin*, F. Utzet** and V. Zaiats***

* National Technical University of Ukraine. Kyëv Polytechnic Institute.

** Universitat Autònoma de Barcelona.

*** Universitat de Vic.

In statistics of stochastic processes and random fields, a moment function or a cumulant of an estimate of either the correlation function or the spectral function can often contain an integral involving a cyclic product of kernels. We define and study this class of integrals and prove a Young-Hölder inequality. This inequality further enables us to study asymptotics of the above mentioned integrals in the situation where the kernels depend on a parameter. An application to the problem of estimation of the response function in a Volterra system is given.

Keywords: Integral involving a cyclic product of kernels, cumulant, Young-Hölder inequality, cross-correlogram, asymptotic normality.

○○○

SORT, vol. 28, 1, pp. 9-26, 2004

SE

Asymptotic normality of the integrated square error of a density estimator in the convolution model

C. Butucea

Modal'X, Université Paris X.

In this paper we consider a kernel estimator of a density in a convolution model and give a central limit theorem for its integrated square error (ISE). The kernel estimator is rather classical in minimax theory when the underlying density is recovered from noisy observations. The kernel is xed and depends heavily on the distribution of the noise, supposed entirely known. The bandwidth is not xed, the results hold for any sequence of bandwidths decreasing to 0. In particular the central limit theorem holds for the bandwidth minimizing the mean integrated square error (MISE). Rates of convergence are sensibly dierent in the case of regular noiseand of super-regular noise. The smoothness of the underlying unknown density is relevant for the evaluation of the MISE.

Keywords: Convolution density estimation, Nonparametric density estimation, Central Limit Theorem, Integrated Squared Error, Noisy observations.

○○○

QÜESTIÓ, vol. 22, 1, pp. 103-116, 1998

NO

Implementación de un algoritmo primal-dual de orden superior mediante el uso de un método predictor-corrector para programación lineal***An implementation of a higher-order primal-dual interior point algorithm using a predictor-corrector method for linear programming***

J. Castro

Estadística i Investigació Operativa. Dept. d'Enginyeria Química. Universitat Rovira i Virgili.

Se presenta una implementación de un algoritmo primal-dual de punto interior para la solución de problemas lineales. El algoritmo difiere de otros ya existentes (como el implementado en el sistema LoQo) en el hecho de que soluciona las denominadas "ecuaciones normales en forma primal" (LoQo soluciona el denominado "sistema aumentado") y en que realiza una clara distinción entre variables acotadas superior e inferiormente, y aquellas sólo acotadas inferiormente. La eficiencia de la implementación es comparada con el sistema LoQo. Para la comparación se utilizan 80 problemas lineales de la colección Netlib (Gay (1985)) (una batería estándar de problemas de programación lineal). Este trabajo es el primero de una serie de dos, cuyo objetivo es la resolución eficiente de problemas cuadráticos por técnicas de punto interior.

Palabras clave: Algoritmo primal-dual, métodos de punto interior, método predictor-corrector, programación lineal.

○○○

SORT, vol. 31, 2, pp. 181-200, 2007

BS

Bayesian joint modelling of the mean and covariance structures for normal longitudinal data

E. Cepeda-Cuervo* and V. Núñez-Antón**

* Universidad Nacional de Colombia.

** Universidad del País Vasco.

We consider the joint modelling of the mean and covariance structures for the general antedependence model, estimating their parameters and the innovation variances in a longitudinal data context. We propose a new and computationally efficient classic estimation method based on the Fisher scoring algorithm to obtain the maximum likelihood estimates of the parameters. In addition, we also propose a new and innovative Bayesian methodology based on the Gibbs sampling, properly adapted for longitudinal data analysis, a methodology that considers linear mean structures and unrestricted covariance structures for normal longitudinal data. We illustrate the proposed methodology and study its strengths and weaknesses by analyzing two examples, the race and the cattle data sets.

Keywords: Antedependence models, Bayes estimation, Fisher scoring, Gibbs sampling.

○○○

SORT, vol. 29, 1, pp. 27-42, 2005

MS

Correspondence analysis and two-way clustering

A. Ciampi*, A. González Marcos** and M. Castejón Limas**

* McGill University, Montreal, Canada.

** University of León.

Correspondence analysis followed by clustering of both rows and columns of a data matrix is proposed as an approach to two-way clustering. The novelty of this contribution consists of: i) proposing a simple method for the selecting of the number of axes; ii) visualizing the data matrix as is done in micro-array analysis; iii) enhancing this representation by emphasizing those variables and those individuals which are 'well represented' in the subspace of the chosen axes. The approach is applied to a 'traditional' clustering problem: the classification of a group of psychiatric patients.

Keywords: Block clustering, selecting number of axes, data visualization.



SORT, vol. 29, 2, pp. 235-248, 2005

TS

On the probability of reaching a barrier in an Erlang(2) risk process

M. M. Claramunt*, M. T. Mármol* and R. Lacayo**

* Universitat de Barcelona.

** Universitat Autònoma de Barcelona.

In this paper the process of aggregated claims in a non-life insurance portfolio as defined in the classical model of risk theory is modified. The Compound Poisson process is replaced with a more general renewal risk process with interoccurrence times of Erlangian type. We focus our analysis on the probability that the process of surplus reaches a certain level before ruin occurs, $\chi(u, b)$. Our main contribution is the generalization obtained in the computation of $\chi(u, b)$ for the case of interoccurrence time between claims distributed as Erlang $(2, \beta)$ and the individual claim amount as Erlang (n, γ) .

Keywords: Risk theory, Erlang distribution, upper barrier, ordinary differential equation, Boundary conditions.



SORT, vol. 27, 1, pp. 1-12, 2003

BI

Likelihood for interval-censored observations from multi-state models

D. Commenges

INSERM Team of Biostatistics.

We consider the mixed discrete-continuous pattern of observation in a multi-state model; this is a classical pattern because very often clinical status is assessed at discrete visit times while time of death is observed exactly. The likelihood can easily be written heuristically for such models. However a formal proof is not easy in such observational patterns. We give a rigorous derivation of the likelihood for the illness-death model based on applying Jacod's formula to an observed bivariate counting process.

Keywords: Multi-state models, illness-death, counting processes, ignorability, interval-censoring, Markov models.

○○○

QÜESTIÓ, vol. 22, 1, pp. 117-142, 1998

NO

Un algoritmo de punto interior para programación cuadrática a través de problemas equivalentes separables

An interior-point algorithm for quadratic programming through separable equivalent problems

J. Castro

Estadística i Investigació Operativa. Dept. d'Enginyeria Química. Universitat Rovira i Virgili.

Se presenta un algoritmo de punto interior para la solución de problemas cuadráticos simétricos y definidos positivos, mediante su transformación en problemas equivalentes separables (esto es, la matriz de coeficientes cuadráticos es diagonal y no existen términos cruzados). El algoritmo difiere de otros ya existentes (como el implementado en el sistema LoQo) en el hecho de que soluciona las denominadas "ecuaciones normales en forma primal" (LoQo soluciona el denominado "sistema aumentado") y en que no requiere ningún tratamiento específico para las variables libres creadas durante la obtención del problema equivalente separable. Se presenta una implementación del algoritmo y su eficiencia es comparada con los sistemas LoQo y Minos 5.3. Para la comparación se utilizan 80 problemas cuadráticos derivados de problemas lineales de la colección Netlib (Gay (1985)) (una batería estándar de problemas de programación lineal). Se obtienen los problemas cuadráticos mediante un generador ad-hoc de problemas cuadráticos.

Palabras clave: Algoritmo primal-dual, métodos de punto interior, método predictor-corrector, programación cuadrática.

○○○

QÜESTIÓ, vol. 24, 1, pp. 151-186, 2000

OS

Avantatges i inconvenients de la metodologia de l'Idescat/INE per elaborar indicadors de la producció industrial per a les regions espanyoles

Advantages and disadvantages of Idescat/INE's methodology to elaborate industrial production indicators for the Spanish regions

M. Clar, R. Ramos i J. Suriñac

Grup de recerca Anàlisi Quantitativa Regional. Universitat de Barcelona.

Conèixer l'evolució conjuntural del sector industrial, tant a nivell nacional com regional, és de gran importància. En aquest sentit, el retard en la publicació de les xifres de les Comptabilitats Nacionals/Regionals, fa necessària l'elaboració d'indicadors que permetin dur a terme un seguiment a curt termini de l'activitat industrial. Així, l'INE elabora un IPI mensual obtingut pel mètode directe pel conjunt de l'Estat. D'altra banda, al llarg dels darrers anys, a algunes comunitats autònomes espanyoles, s'han engegat projectes centrats en l'elaboració d'indicadors de l'activitat industrial regional, tot i que a partir de metodologies no homogènies. Per corregir aquesta situació, d'un temps ençà a diferents fòrums s'ha proposat emprar la metodologia emprada per l'Idescat per elaborar l'indicador de la comunitat catalana com a alternativa per construir indicadors de l'activitat industrial regional, atès el seu bon comportament per a Catalunya. Així, l'INE recentment ha publicat uns IPIs per a les CA espanyoles d'acord amb dita metodologia. A aquest treball s'estudia la idoneïtat d'estendre l'esmentada metodologia a totes les regions espanyoles. Per això, es duu a terme una anàlisi comparativa centrada en tres de les (quatre) regions que disposen d'un IPI elaborat pel mètode directe: Andalusia, Astúries i el País Basc.

Paraules clau: Activitat industrial, índex de producció industrial, indicadors regionals, conjuntura.



QÜESTIÓ, vol. 23, 1, pp. 155-165, 1999

BI

Papel ético del estadístico en la experimentación humana

Statistical ethical issues in human experimentation

E. Cobo

Departament d'Estadística i Investigació Operativa. Secció Informàtica. Universitat Politècnica de Catalunya.

En diferentes ocasiones, desde la Sociedad de Estadística e Investigación Operativa o desde la Región Española de la International Biometry Society, se nos ha pedido a los estadísticos aplicados que hablemos de aquellas características de nuestro trabajo profesional no contempladas en los planes de estudio. En estas líneas pretendemos recordar el relevante papel que el estadístico desempeña en la investigación y desarrollo de nuevas pautas terapéuticas. Para ello, expondremos los principios éticos básicos que guían la experimentación con seres humanos y sus implicaciones en la tarea profesional del estadístico. También esbozaremos aquellas líneas de investigación metodológica básica que podrían aligerar las tensiones entre principios éticos e investigación con seres humanos. Finalmente, insistimos en la necesidad de la reincorporación inmediata de estadísticos en las comisiones encargadas de valorar la validez científica de las propuestas de investigación clínica.

Palabras clave: Ensayos clínicos, ética, código deontológico, estadísticos, causalidad.



SORT, vol. 27, 1, 13-30, 2003

BI

Cumulative processes related to event histories

R. J. Cook, J. F. Lawless and K.-A. Lee
University of Waterloo.

Costs or benefits which accumulate for individuals over time are of interest in many life history processes. Familiar examples include costs of health care for persons with chronic medical conditions, the payments to insured persons during periods of disability, and quality of life which is sometimes used in the evaluation of treatments in terminally ill patients. For convenience, here we use the term costs to refer to cost or other cumulative measures. Two important scenarios are (i) where costs are associated with the occurrence of certain events, so that total cost accumulates as a step function, and (ii) where individuals may move between various states over time, with cost accumulating at a constant rate determined by the state occupied. In both cases, there is frequently a random variable T that represents the duration of the process generating the costs. Here we consider estimation of the mean cumulative cost over a period of interest using methods based upon marginal features of the cost process and intensity based models. Robustness to adaptive censoring is discussed in the context of the multi-state methods. Data from a quality of life study of breast cancer patients are used to illustrate the methods.

Keywords: Adaptive censoring, cumulative processes, marginal methods, multi-state processes, quality adjusted life-years, prevalence function.

○○○

QÜESTIÓ, vol. 24, 1, pp. 85-110, 2000

OS

Diversitat i complementarietat de les fonts estadístiques***Diverseness and complementariness of statistical sources***

À. Costa
Institut d'Estadística de Catalunya.

Es presenta la diversitat de les fonts generals d'estadística oficial, posant de relleu la importància de censos, enquestes, aprofitament de registres administratius i estimacions comptables. Es justifica aquesta diversitat per les diferents prestacions de cada forma de producció, des del punt de vista de la puntualitat i de la quantitat d'informació i també des del punt de vista de la fiabilitat dels resultats. Tot seguit es passa a l'àmbit de les fonts regionals on, de nou, es destaca la diversitat i la complementarietat de les mateixes. En aquest punt destaca la complementarietat en els processos de producció d'informació estadística i en els resultats finals. Pel que fa als processos, es presenta una tipologia de formes de col·laboració entre l'administració central (INE) i l'autonòmica (Idescat). En la complementarietat dels resultats es tornen a fer servir els mateixos conceptes emprats per a les fonts generals: puntualitat, quantitat d'informació i fiabilitat.

Paraules clau: Fonts estadístiques, diversitat de fonts, complementarietat de fonts, correspondència, coherència, territori/temps/conceptes, col·laboració institucional.

○○○

QÜESTIÓ, vol. 26, 1-2, pp. 213-243, 2002

OS

Estimadores compuestos en estadística regional: una aplicación a la estimación de la tasa de variación de la ocupación en la industria

Composite estimators in regional statistics: an application to the estimation of the rate of change in industrial occupation

À. Costa*, A. Satorra** y E. Ventura**

* Institut d'Estadística de Catalunya.

** Departament d'Economia i Empresa. Universitat Pompeu Fabra.

Este trabajo es parte de un proyecto que estudia la aplicación de estimadores compuestos (combinación de estimadores directos e indirectos) para áreas pequeñas en estadística regional. Comparamos tres estimadores: uno directo basado en datos muestrales de cada Comunidad Autónoma (CA), otro sintético (indirecto) que combina los datos estatales con información específica de las CCAA, y un tercer estimador, el compuesto, basado en un modelo estadístico que se concreta en una combinación lineal de los dos anteriores. Ilustramos el método de análisis adoptado mediante la estimación de la tasa de variación de la ocupación industrial en las diferentes CCAA de España.

Palabras clave: Estadística regional, áreas pequeñas, estimadores compuestos.

○○○

SORT, vol. 27, 1, pp. 113-136, 2003

OS

An empirical evaluation of small area estimators

À. Costa*, A. Satorra** and E. Ventura**

* Institut d'Estadística de Catalunya.

** Departament d'Economia i Empresa. Universitat Pompeu Fabra.

This paper compares five small area estimators. We use Monte Carlo simulation in the context of both an artificial and real populations. In addition to the direct and indirect estimators, we consider the optimal composite estimator with population weights, and two composite estimators with estimated weights: one that assumes homogeneity of within area variance and squared bias and one that uses area-specific estimates of variance and squared bias. In the study with real population, we found that among the feasible estimators, the best choice is the one that uses area-specific estimates of variance and squared bias.

Keywords: Regional statistics, small areas, root mean square error, direct, indirect and composite estimators.

○○○

SORT, vol. 28, 1, pp. 69-86, 2004

OS

Improving both domain and total area estimation by composition

À. Costa*, A. Satorra** and E. Ventura**

* Institut d'Estadística de Catalunya.

** Department of Economics and Business. Universitat Pompeu Fabra.

In this article we propose small area estimators for both the small and large area parameters. When the objective is to estimate parameters at both levels, optimality is achieved by a sample design that combines fixed and proportional allocation. In such a design, one fraction of the sample is distributed proportionally among the small areas and the rest is evenly distributed. Simulation is used to assess the performance of the direct estimator and two composite small area estimators, for a range of sample sizes and different sample distributions. Performance is measured in terms of mean squared errors for both small and large area parameters. Small area composite estimators open the possibility of reducing the sample size when the desired precision is given, or improving precision for a given sample size.

Keywords: Regional statistics, small areas, mean square error, direct and composite estimators.

○○○

SORT, vol. 30, 1, pp. 101-122, 2006

OS

Improving small area estimation by combining surveys: new perspectives in regional statistics

À. Costa*, A. Satorra** and E. Ventura**

* Institut d'Estadística de Catalunya.

** Department of Economics and Business. Universitat Pompeu Fabra.

A national survey designed for estimating a specific population quantity is sometimes used for estimation of this quantity also for a small area, such as a province. Budget constraints do not allow a greater sample size for the small area, and so other means of improving estimation have to be devised. We investigate such methods and assess them by a Monte Carlo study. We explore how a complementary survey can be exploited in small area estimation. We use the context of the Spanish Labour Force Survey (EPA) and the Barometer in Spain for our study.

Keywords: Composite estimator, complementary survey, mean squared error, official statistics, regional statistics, small areas.

○○○

SORT, vol. 30, 1, pp. 55-84, 2006

PD

The importance of being the upper bound in the bivariate family

C. M. Cuadras
University of Barcelona.

Any bivariate cdf is bounded by the Fréchet-Hoeffding lower and upper bounds. We illustrate the importance of the upper bound in several ways. Any bivariate distribution can be written in terms of this bound, which is implicit in logit analysis and the Lorenz curve, and can be used in goodness-of-fit assesment. Any random variable can be expanded in terms of some functions related to this bound. The Bayes approach in comparing two proportions can be presented as the problem of choosing a parametric prior distribution which puts mass on the null hypothesis. Accepting this hypothesis is equivalent to reaching the upper bound. We also present some parametric families making emphasis on this bound.

Keywords: Hoeffding's lemma, Fréchet-Hoeffding bounds, given marginals, diagonal expansion, logit analysis, goodness-of-fit, Lorenz curve, Bayes test in 2×2 tables.



SORT, vol. 29, 1, 57-100, 2005

OS

Automatic error localisation for categorical, continuous and integer data pp

T. de Waal
Statistics Netherlands.

Data collected by statistical offices generally contain errors, which have to be corrected before reliable data can be published. This correction process is referred to as statistical data editing. At statistical offices, certain rules, so-called edits, are often used during the editing process to determine whether a record is consistent or not. Inconsistent records are considered to contain errors, while consistent records are considered error-free. In this article we focus on automatic error localisation based on the Fellegi-Holt paradigm, which says that the data should be made to satisfy all edits by changing the fewest possible number of fields. Adoption of this paradigm leads to a mathematical optimisation problem. We propose an algorithm for solving this optimisation problem for a mix of categorical, continuous and integer-valued data. We also propose a heuristic procedure based on the exact algorithm. For five realistic data sets involving only integer-valued variables we evaluate the performance of this heuristic procedure.

Keywords: Branch-and-bound, categorical data, continuous data, error localisation, Fourier-Motzkin elimination, integer-valued data, statistical data editing.



QÜESTIÓ, vol. 25, 1, pp. 19-46, 2001

TS

Comportamiento de los contrastes ADF, PP y KPSS al trabajar con series ajustadas de estacionalidad***ADF, PP & KPSS tests performance when applied to seasonal adjusted time series***

T. del Barrio, A. del Sur y J. Suriñach

Departamento de Econometría, Estadística y E.E. Grup de Recerca de Qualitat "Anàlisi Quantitativa Regional". Facultat de Econòmiques. Universitat de Barcelona.

En este trabajo se analiza el comportamiento de los tests de raíces unitarias cuando se utilizan los componentes ciclo-tendencia obtenidos a partir de procedimientos de extracción de señales en lugar de utilizar las series originales. Adicionalmente se intenta detectar las causas finales de los efectos perniciosos observados. Los procedimientos de extracción de señales analizados son el basado en modelos ARIMA y el filtro de líneas aéreas modificado. Un ejercicio de simulación nos permite concluir que se puede llegar a detectar un orden de integrabilidad superior en la frecuencia cero para las series filtradas que para las series originales. La causa principal de los resultados obtenidos se encuentra en el cumplimiento del requisito canónico por los filtros de los procedimientos analizados.

Palabras clave: Contrastes de integrabilidad y estacionariedad, extracción de señales.



QÜESTIÓ, vol. 26, 3, pp. 483-501, 2002

MA

Anàlisi de matrius quadrades no simètriques: un enfocament integral usant anàlisi de correspondències***Analysis of square skew-symmetric tables: An integral vision by correspondence analysis***

J. Daunis*, T. Aluja** i S. Thió*

* Departament d'Informàtica i Matemàtica Aplicada. Universitat de Girona.

** Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya.

En aquest article s'introdueix una proposta integrada d'anàlisi de matrius quadrades no simètriques mitjançant anàlisi de correspondències. En ella s'han integrat les dues principals famílies de metodologies que proposen solucions a la problemàtica inherent d'aquest tipus de taules: les diagonals sobrecarregades i els zeros estructurals. S'aplica a l'exemple de l'estudi dels moviments de població deguts al treball entre les 41 comarques catalanes.

Paraules clau: Matrius quadrades, diagonal sobrecarregada, anàlisi de correspondències, zeros estructurals, descomposició en valors singulars, dades mancants.



SORT, vol. 29, 2, pp. 269-288, 2005

BI

A comparison of parametric models for mortality graduation. Application to mortality data of the Valencia Region

A. Debón*, F. Montes** and R. Sala**

* Universidad Politécnica de Valencia.

** Universitat de València.

The parametric graduation of mortality data has as its objective the satisfactory estimation of the death rates based on mortality data but using an age-dependent function whose parameters are adjusted from the crude rates obtainable directly from the data. This paper proposes a revision of the most commonly used parametric methods and compares the results obtained with each of them when they are applied to the mortality data for the Valencia Region. As a result of the comparison, we conclude that the Gompertz-Makeham functions estimated by means of generalized linear models lead to the best results. Our working method is of additional interest for being applicable to mortality data for a wide range of ages from any geographical conditions, allowing us to select the most appropriate life table for the case in hand.

Keywords: Gompertz-Makeham functions, Heligman and Pollard's laws, parametric graduation.



QÜESTIÓ, vol. 26, 3, pp. 537-559, 2002

AO

Problema de contratación de carretilleros para un almacén de productos manufacturados

Labor requirements for vehicle loading problem in a store of manufactured products

C. R. Delgado, S. Casadó y J. F. Alegre

Facultad C. Económicas y Empresariales. Universidad de Burgos.

En este trabajo se analiza un problema planteado recientemente a sus autores por una empresa fabricante de componentes de automóviles. Dicha empresa almacena sus productos manufacturados hasta que los clientes (compradores) pasan a recogerlos. Los clientes solicitan sus productos con una frecuencia conocida. Se trata de determinar en función de dichas frecuencias, en qué fechas y a que horas o slots, han de pasar los clientes a recoger sus pedidos. Fijado el horizonte temporal objeto de estudio, el objetivo para la empresa es minimizar en ese conjunto de días, el número de carretilleros necesarios para cargar los pedidos en los camiones de los clientes. El número de carretilleros diario viene determinado por el slot más ocupado. En este problema se han de tomar decisiones a dos niveles: elaboración del calendario de entrega para cada uno de los pedidos, y asignación diaria de pedidos a slots. No se ha encontrado en la literatura ningún trabajo que estudie o pueda ajustarse a este problema. Se diseñan y analizan 3 sencillos Metaheurísticos: uno sigue un proceso de Búsqueda Tabú básico, otro es un procedimiento de Búsqueda en Entorno Variable y el tercero es un Algoritmo Evolutivo. Se realizan pruebas con instancias ficticias y por último se resuelven las instancias reales planteadas a los autores de este trabajo.

Palabras clave: Logística, Multiprocessor Scheduling Problem (MSP), Búsqueda Local, Búsqueda Tabú, Búsqueda en Entorno Variable, Algoritmos Evolutivos.

○○○

SORT, vol. 29, 1, pp. 101-118, 2005

PD

Estimation of the spectral density of a homogeneous random stable discrete time field

N. N. Demesh and S. L. Chekhmenok
Belarus State University.

In earlier papers, 2π -periodic spectral data windows have been used in spectral estimation of discrete-time random fields having finite second-order moments. In this paper, we show that 2π -periodic spectral windows can also be used to construct estimates of the spectral density of a homogeneous symmetric α -stable discrete-time random field. These fields do not have second-order moments if $0 < \alpha < 2$. We construct an estimate of the spectrum, calculate the asymptotic mean and variance, and prove weak consistency of our estimate.

Keywords: Homogeneous stable fields, spectral density estimate.

○○○

QÜESTIÓ, vol. 25, 2, pp. 287-300, 2001

AO

Modelización de un DSS para la gestión de productos perecederos

DSS modelling for perishable inventory management

B. Díaz Fernández, J. A. del Brío González y B. González Torre
ETS Ingenieros Industriales. Universidad de Oviedo.

La gestión de inventarios de productos perecederos ha atraído desde hace tiempo la atención de los investigadores de Dirección de Operaciones. En este artículo se presenta la modelización e implementación de un sistema de apoyo a la toma de decisiones (DSS) para la gestión de productos perecederos, aplicado a la distribución interhospitalaria de hemoderivados. En estos casos se trata de satisfacer en lo posible las demandas, tratando de evitar a la vez la caducidad de los productos en manos de los clientes (los cuales teóricamente no sufren coste extra por sobrestock). Dada la naturaleza multicriterio del problema, para la modelización se ha usado goal programming, con resultados que en casos concretos permiten reducciones considerables en la cantidad de productos inutilizables.

Palabras clave: Sistemas de apoyo a la toma de decisiones, programación por metas, gestión de inventarios.

○○○

QÜESTIIÓ, vol. 22, 1, 3-37, 1998

RE

El modelo lineal sin término independiente y el coeficiente de determinación. Un estudio Monte Carlo***The linear model without a constant term and the coefficient of determination. A Monte Carlo study***

R. Dios Palomares

Dpto. de Estadística e Investigación Operativa. Escuela Técnica Superior de Ingenieros Agrónomos y de Montes de la Universidad de Córdoba.

En el presente trabajo se analiza y compara mediante un experimento Monte Carlo el comportamiento de cinco expresiones para el Coeficiente de Determinación cuando el modelo lineal se especifica sin término independiente. Se ensayan distintos valores del parámetro poblacional P_2 , que mide la proporción de varianza explicada por el modelo, introduciendo también la multicolinealidad como factor de variación en el diseño. Se confirma el coeficiente propuesto por Heijmans y Neudecker (1987) y el de Barten (1987), como idóneos para medir la bondad del modelo.

Palabras clave: Coeficiente de determinación, bondad de ajuste, modelo lineal sin término independiente, método Monte Carlo.

○○○

QÜESTIIÓ, vol. 23, 3, pp. 437-464, 1999

RE

Análisis comparativo de estimadores pretest de heterocedasticidad en modelos econométricos. Un estudio Monte Carlo***A comparative analysis of heterocedasticity pretest estimators in econometrics models. A Monte Carlo study***

R. Dios y C. Rodríguez Fonseca

Universidad de Córdoba. Departamento de Estadística, Econometría, Investigación Operativa y Organización de Empresas.

En el presente artículo se recogen los resultados de una investigación llevada a cabo sobre el comportamiento de pretest de heterocedasticidad. Con este fin se ha diseñado un experimento Monte Carlo, introduciendo como proceso generador de datos un modelo con tres supuestos sobre la estructura de la varianza del error y con distintos niveles de heterocedasticidad para cada uno de ellos. Asimismo, se analiza la potencia de los diferentes contrastes de heterocedasticidad bajo los distintos supuestos de estructura heterocedástica. Además, se estudia la consecuencia de utilizar un estimador por Mínimos Cuadrados Generalizados Factibles cuya matriz $\hat{\Omega}$ no se corresponde con la matriz Ω del proceso generador de datos, mediante una función de riesgo de los estimadores pretest. Se concluye que es más importante el procedimiento de estimación empleado que el contraste aplicado para detectar la heterocedasticidad.

Palabras clave: Test de heterocedasticidad, pretest, Monte Carlo, curva de potencia.

○○○

QÜESTIÓ, vol. 26, 3, pp. 443-459, 2002

SE

Analysis on the individual efficiency prediction in the composed error frontier model. A Monte Carlo Study

R. Dios-Palomares, A. Ramos Millán and J. A. Roldán-Casas
Department of Statistics E.T.S.I.A.M. University of Córdoba.

This study seeks to analyse some important questions related to the Stochastic Frontier Model, such as the method proposed by Jondrow et al (1982) to separate the error term into its two components, and the measure of efficiency given by Timmer (1971). To this purpose, a Monte Carlo experiment has been carried out using the Half-Normal and Normal-Exponential specifications throughout the rank of the γ parameter. The estimation errors have been eliminated, so that the intrinsic variability of the conditional of u given ε can be evaluated. In addition, the behaviour of the mean and mode as point estimators of u is investigated. The results have yielded some interesting findings. We have observed that both the point estimates and the mean efficiency are more precise in cases of lower efficiency. This occurs when the variable that generates the inefficiency outweighs the one that picks up the errors out of the control. The change in order found between the estimated efficiency and its true value is misleadingly high especially for low γ , which underlines the risk of estimating at these values.

Keywords: Bias, MSE, point estimator, production frontier, Jondrow formulae.



QÜESTIÓ, vol. 22, 3, pp. 493-510, 1998

OS

La verificació aleatòria: una estratègia per millorar i avaluar la qualitat de l'entrada de dades***Random verification: a strategy to improve and assess the quality of data entry***

J. M. Domènech Massons, J. M. Losilla Vidal i M. Potell Vidal
Dept. de Psicobiologia i de Metodologia de les Ciències de la Salut. Universitat Autònoma de Barcelona.

S'aborda la problemàtica de la reducció dels errors que es produeixen durant la introducció de les dades i que no es poden controlar mitjançant proteccions automàtiques. Enfront aquest problema, l'estratègia habitual és la "doble entrada" (DE) de les dades, la qual augmenta considerablement el cost de la investigació. Com alternativa a aquesta estratègia es proposa un nou procediment, implementat en el Sistema DAT, que es basa en un procés de "verificació aleatòria" (VA) d'un percentatge del total de dades. A més de reduir el cost, la VA ofereix altres avantatges com el fet de proporcionar una estimació del percentatge d'errors, i oferir un índex d'aptitud i eficàcia dels operadors. A la segona part de l'article es presenten els resultats d'un experiment que recolza la hipòtesi que la VA augmenta l'eficàcia de l'entrada de dades, sense minva de l'eficiència, quan es compara amb situacions en les quals no s'aplica cap control que permeti obtenir un indicador sobre la qualitat de les dades que s'introdueixen.

Paraules clau: Qualitat de les dades, doble entrada, entrada de dades, verificació aleatòria, Sistema DAT.



QÜESTIÓ, vol. 23, 2, pp. 225-237, 1999

ID

Some results involving the concepts of moment generating function and affinity between distribution functions. Extension for r k -dimensional normal distribution functions

A. Dorival Campos

Dpto. de Matemàtica Aplicada à Biologia. Faculdade de Medicina de Riberirao Preto. Universidade de Sao Paulo. Brasil.

We present a function $\rho(F_1, F_2, t)$ which contains Matusita's affinity and express the "affinity" between moment generating functions. An interesting result is expressed through decomposition of this "affinity" $\rho(F_1, F_2, t)$ when the functions considered are k -dimensional normal distributions. The same decomposition remains true for others families of distribution functions. Generalizations of these results are also presented.

Keywords: Affinity, moment generating functions, distance, inner product, multivariate normal distributions, probability density functions, absolutely convergent.



SORT, vol. 29, 2, pp. 249-268, 2005

SE

Factorial experimental designs and generalized linear models

S. Dossou-Gbété and W. Tinsson

Université de Pau et des Pays de l'Adour.

This paper deals with experimental designs adapted to a generalized linear model. We introduce a special link function for which the orthogonality of design matrix obtained under Gaussian assumption is preserved. We investigate by simulation some of its properties.

Keywords: Generalized linear model, exponential family, Fisher-Scoring algorithm, factorial designs, regular fraction.



QÜESTIÓ, vol. 23, 1, pp. 113-128, 1999

OS

Le traitement de l'incertitude dans les perspectives démographiques

Processing uncertainty in demographic projections

J. Duchêne

Institut de Démographie, Université Catholique de Louvain, Belgique.

Les tendances prédites sont souvent des extrapolations des tendances passées observées jusqu'à la date d'établissement des projections alors que la réalité peut présenter des retournements de tendances ou des irrégularités inattendues. L'incertitude étant comprise dans le sens d'incapacité à prévoir l'avenir avec exactitude, les erreurs de prévision sont la manifestation de cette incertitude et proviennent de l'imprécision dont l'extrapolation des composantes du changement démographique est affectée et qui s'accroît lorsqu'on s'éloigne de la date de début de la période de projection. Les erreurs ont tendance à croître avec la longueur de l'horizon temporel de la projection. Presque tous les auteurs de prévisions démographiques officielles construisent plusieurs séries basées sur la combinaison de scénarios. Ils ne fournissent toutefois pas la probabilité avec laquelle la réalité se situera dans l'intervalle délimité par les variantes haute et basse. Cet article dresse une liste de quelques mesures utilisées pour évaluer la précision des perspectives et précise leurs limites et avantages respectifs. Il est par ailleurs proposé d'examiner les erreurs exposées par une méthode qui s'apparente à la décomposition des effets âge-période-cohorte ou d'ajuster un modèle de série chronologique aux erreurs de projection observées sur certaines variables dans les projections antérieures afin d'établir un intervalle de confiance pour ces variables à appliquer à la nouvelle projection. Il est enfin proposé de raisonner sur la base de modèles stochastiques ou d'utiliser une méthode de projections probabilistes de population basée sur des opinions d'experts à propos des tendances futures de la fécondité, de la mortalité et de la mobilité spatiale ainsi qu'à propos du niveau d'incertitude dont ces tendances peuvent être entachées.

Mots clés: Incertitude, projection démographique.



QÜESTIÓ, vol. 25, 3, pp. 437-477, 2001

SE

Estimación no paramétrica de la función de riesgo: aplicaciones a sismología

Nonparametric estimation of the hazard function: applications to seismology

G. Estévez y A. Quintela

Departamento de Matemáticas-Facultad de Informática. Universidad de A Coruña.

Se estudia la estimación de tipo no paramétrico de la función de riesgo o razón de fallo de una variable aleatoria real. A partir de una muestra X_1, X_2, \dots, X_n de datos no censurados y no necesariamente independientes, se considera un estimador cociente entre el estimador núcleo de la función de densidad y un estimador núcleo de la función de supervivencia, sobre el que se estudia el problema de selección del parámetro ventana. Por medio de un estudio de simulación se observa la ventaja de utilizar este estimador frente al que estima la función de supervivencia a través de la función de distribución empírica. Finalmente, se realiza una aplicación práctica a datos de terremotos sucedidos en California y en Granada.

Palabras clave: Estimación no paramétrica, procesos fuertemente mixing, validación cruzada, razón de fallo.



QÜESTIÓ, vol. 23, 1, pp. 129-152, 1999

OS

Mètodes i models per projectar els components del canvi demogràfic en les projeccions de població de l'Institut d'Estadística de Catalunya

Methods and models to project the components of demographic change in the population projections of the Institut d'Estadística de Catalunya

M. Farré Mallofré i J. A. Sánchez Cepeda
Institut d'Estadística de Catalunya.

Aquest treball descriu la realització de les projeccions de població 1996-2010-2030 de l'Institut d'Estadística de Catalunya (Idescat) des del punt de vista metodològic. Es descriuen les metodologies particulars desenvolupades per a cada un dels components del creixement demogràfic. Aquestes metodologies no només se centren en els valors dels indicadors demogràfics a projectar sinó que introdueixen comportaments específics i dinàmics per als diferents grups d'edat de la població. En la fecunditat s'ha aplicat un mètode que integra la descendència final de les diferents generacions de dones, l'edat a la maternitat i la descendència segons ordre de naixement. En la mortalitat s'ha aplicat un mètode que permet combinar l'evolució de l'esperança de vida (traç cronològic) amb evolucions no homogènies de les taxes de mortalitat a diferents edats (coeficients de millora). En la migració s'ha treballat amb el model migratori de Rogers, aplicant uns perfils dinàmics en el temps. Finalment, es fa una anàlisi de la incertesa inherent als resultats i la seva quantificació.

Paraules clau: Projeccions de població, fecunditat, mortalitat, migració, descendència final, traç cronològic, perfil migratori.



QÜESTIÓ, vol. 23, 2, pp. 263-296, 1999

OS

Serveis estadístics. Preparant-se per al futur

Statistical services. Preparing for the future

I. Fellegi
Statistics Canada.

L'objectiu d'aquesta anàlisi és mirar endavant: cap als reptes als quals s'enfronten les agències estadístiques i com els han d'encarar. S'analitza el context dins el qual han d'evolucionar els instituts d'estadística: les principals forces que modifiquen l'economia i la societat, les mesures polítiques emergents que aparentment requereixen informació estadística nova o addicional i els canvis en la naturalesa dels governs que tenen repercussions significatives per als instituts d'estadística. El gruix de l'article proposa una "estratègia interna" per als instituts d'estadística. Això inclou el desenvolupament de nous sistemes d'informació estadística que vagin més enllà de la tradicional funció d'"observació" dels fenòmens i intentin aportar informació sobre les seves causes. Un altre component clau d'aquesta estratègia ha de ser l'assoliment d'un alt nivell d'adaptabilitat, la qual exigeix iniciatives específiques. Finalment, l'article explora els elements d'una "estratègia externa": l'assoliment i el manteniment d'un alt nivell de pertinència; l'objectivitat política i la seva percepció, i la col·laboració internacional.

Paraules clau: Estadística oficial, reptes polítics, sistemes d'informació, flexibilitat, pertinència, col·laboració internacional.

○○○

SORT, vol. 28, 1, pp.55-68, 2004

TS

Modelling stock returns with AR-GARCH processes

E. Ferenstein^{*,**} and M. Gąsowski^{***}

* Faculty of Mathematics and Information Science. Warsaw University of Technology. Poland.

** Polish-Japanese Institute of Information Technologies. Warsaw, Poland.

*** Bank Gospodarki Żywnościowej S.A.

Financial returns are often modeled as autoregressive time series with random disturbances having conditional heteroscedastic variances, especially with GARCH type processes. GARCH processes have been intensely studying in financial and econometric literature as risk models of many financial time series. Analyzing two data sets of stock prices we try to fit AR (1) processes with GARCH or EGARCH errors to the log returns. Moreover, hyperbolic or generalized error distributions occur to be good models of white noise distributions.

Keywords: Autoregressive process, GARCH and EGARCH models, conditional heteroscedastic variance, financial log returns.

○○○

QÜESTIÓ, vol. 25, 1, pp. 47-68, 2001

GS

Concatenación temporal de modelos espaciales y su aplicación al estudio de la meningitis en España

Temporal concatenation of spatial models and its application to the study of meningitis in Spain

J. Ferrándiz*, F. Martínez** y P. Sanmartín***

* Departament d'Estadística i I.O., Universitat de València.

** Instituto de Salud Carlos III, Madrid.

*** Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena.

La cartografía de enfermedades infecciosas en periodos sucesivos plantea la necesidad de su extensión al caso dinámico. En este trabajo proponemos la concatenación temporal de modelos auto-regresivos espaciales para abordar el análisis de mortalidad por meningitis en España en el período 1950-1990 con datos agregados a nivel provincial. Para la estimación y selección del modelo usamos técnicas basadas en la función de verosimilitud.

Palabras clave: Análisis espacio-temporal, modelos espaciales auto-regresivos, enfermedades infecciosas.

○○○

SORT, vol. 27, 2, pp. 165-174, 2003

MA

Asymptotic study of canonical correlation analysis: from matrix and analytic approach to operator and tensor approach

J. Fine

Laboratoire de Statistique et Probabilités, Université Paul Sabatier. France.

Asymptotic study of Canonical Correlation Analysis gives the opportunity to present the different steps of an asymptotic study and to show the interest of an operator and tensor approach of multidimensional asymptotic statistics rather than the classical, matrix and analytic approach. Using the last approach, Anderson (1999) assumes the random vectors to have a normal distribution and the non zero canonical correlation coefficients (c.c.c.) to be distinct. The new approach we use, Fine (2000), is free coordinate, free distribution and permits to have no restriction on the c.c.c. multiplicity order. Of course, when vectors have a normal distribution and when the non zero c.c.c. are distinct, it is possible to find again Anderson's results but we diverge on two of them. In this methodological presentation, we insist on the analysis frame (Dauxois and Pousse, 1976), the sampling model (Dauxois, Fine and Pousse, 1979) and the different mathematical tools (Fine, 1987, Dauxois, Romain and Viguer, 1994) which permit to solve problems encountered in this type of study, and even to obtain asymptotic behavior of the analyses random elements (principal components, canonical variables, ...).

Keywords: Multivariate analysis, canonical correlation analysis, asymptotic study, operator, free coordinate, free distribution.

○○○

QÜESTIÓ, vol. 23, 1, pp. 3-39, 1999

ST

Una revisión de diferentes aportaciones al diseño en poblaciones finitas***A comparative review of different contributions in survey sampling***

B. Font

Universitat de València.

Este artículo es una revisión de los resultados más relevantes de diseño en poblaciones finitas. Los resultados presentados se clasifican en tres grupos: población fija, inferencia basada en modelos de superpoblación clásicos e inferencia basada en modelos de superpoblación Bayesianos, y se analizan de forma comparativa. Entre los resultados revisados señalemos las aportaciones sobre: procedimiento de muestreo, tamaño óptimo de la muestra y procedimientos de estratificación.

Palabras clave: Diseño óptimo, estrategias de muestreo, inferencia Bayesiana, modelos de superpoblación Bayesianos, modelos de superpoblación clásicos, población fija.

○○○

QÜESTIÓ, vol. 23, 1, pp. 43-60, 1999

GT

Amplitude of weighted majority game strict representations

J. Freixas and A. Puente

Department of Applied Mathematics III. Polytechnic University of Catalonia. Manresa.

Some real situations which may be described as weighted majority games can be modified when some players increase or decrease their weights and/or the quota is modified. Nevertheless, some of these modifications do not change the game. In the present work we shall estimate the maximal percentage variations in the weights and the quota which may be allowed without changing the game (amplitude). For this purpose, we have to use strict representations of weighted majority games.

Keywords: Simple games, weighted majority games, strict representations, linearly separable function, tolerance, amplitude.

○○○

SORT, vol. 30, 2, pp. 125-170, 2006

SE

What does intrinsic mean in statistical estimation?

G. García* and J. M. Oller**

* Pompeu Fabra University.

** University of Barcelona.

In this paper we review different meanings of the word intrinsic in statistical estimation, focusing our attention on the use of this word in the analysis of the properties of an estimator. We review the intrinsic versions of the bias and the mean square error and results analogous to the Cramér-Rao inequality and Rao-Blackwell theorem. Different results related to the Bernoulli and normal distributions are also considered.

Keywords: Intrinsic bias, mean square Rao distance, information metric.

○○○

QÜESTIÓ, vol. 25, 3, pp. 499-508, 2001

RE

Aplicaciones empresariales de Data Mining***Data Mining business applications***

L. Garrido y J. I. Latorre

Departament d'Estructura i Constituents de la Matèria. Universitat de Barcelona.

El Data Mining, o extracción de información útil y no evidente de grandes bases de datos, es una tecnología con un gran potencial para ayudar a las empresas a focalizar sus esfuerzos alrededor de la información importante contenida en sus "data warehouses". En este

artículo analizaremos las ideas básicas que sustentan el Data Mining y, más concretamente, la utilización de redes neuronales como herramienta estadística avanzada. Presentaremos también dos ejemplos reales de la aplicación de estas técnicas: predicciones bursátiles y predicción de propagación de fuego en cables eléctricos.

Palabras clave: Minería de datos, data mining, redes neuronales.



SORT, vol. 27, 1, pp. 31-40, 2003

BI

A sensitivity analysis for causal parameters in structural proportional hazards models

E. Goetghebeur and T. Loeys
Ghent University.

Deviations from assigned treatment occur often in clinical trials. In such a setting, the traditional intent-to-treat analysis does not measure biological efficacy but rather programmatic effectiveness. For all-or-nothing compliance situation, Loeys and Goetghebeur (2003) recently proposed a Structural Proportional Hazards method. It allows for causal estimation in the complier subpopulation provided the exclusion restriction holds: randomization per se has no effect unless exposure has changed. This assumption is typically made with structural models for noncompliance but questioned when the trial is not blinded. In this paper we extend the structural PH model to allow for an effect of randomization per se. This enables analyzing sensitivity of conclusions to deviations from the exclusion restriction. In a colorectal cancer trial we find the causal estimator of the effect of an arterial device implantation to be remarkably insensitive to such deviations.

Keywords: Causal inference, compliance, exclusion restriction, proportional hazards, randomized clinical trials, sensitivity análisis.



QÜESTIÓ, vol. 23, 2, pp. 365-392, 1999

BI

Estudios de supervivencia con datos no observados. Dificultades inherentes al enfoque paramétrico

Survival studies with non observed data. Difficulties concerning the parametric approach

G. Gómez* y C. Serrat**

* Departament d'Estadística i Investigació Operativa. Universitat Politècnica de Catalunya.

** Departament de Matemàtica Aplicada I. Universitat Politècnica de Catalunya.

A partir de una muestra de datos de supervivencia que contiene valores no observados en las covariantes de interés, presentamos una metodología que permite extraer toda la información contenida en covariantes completamente observadas, que estén fuertemente correlacionadas con las citadas covariantes de interés. El enfoque utilizado es completamente

paramétrico y se basa en el método de máxima verosimilitud. Mostramos las dificultades, tanto de índole práctica como filosófica, que aparecen en la especificación de la función de verosimilitud y en su optimización. Diseñamos una metodología que permite determinar en qué medida los estimadores resultantes dependen de la modelización del patrón de no respuesta y la implementamos sobre S-PLUS. Dicha metodología, a su vez, permite el estudio de la tipología de datos no observados y proporciona un análisis de sensibilidad de los resultados obtenidos. Ilustramos la metodología presentada y sus dificultades con una aplicación a una cohorte de pacientes con tuberculosis pulmonar infectados por el virus de la inmunodeficiencia humana.

Palabras clave: Análisis de supervivencia, estudio de validación, máxima verosimilitud, modelos de datos incompletos, patrón de no respuesta aleatorio, patrón de no respuesta no ignorable.



QÜESTIÓ, vol. 25, 3, pp. 393-414, 2001

PD

Una generalización de los procesos estocásticos log-normal y de Gompertz como procesos de Itô

A generalization of the log-normal and Gompertz stochastic processes as Itô processes

J. Gómez García y F. Buendía Moya

Departamento de Métodos Cuantitativos para la Economía. Facultad de Economía y Empresa. Universidad de Murcia.

Estudiamos una ecuación diferencial estocástica de Itô que es una generalización de los modelos estocásticos logarítmico-normal y de Gompert. Reducimos la ecuación mediante una transformación de cambio de estado a otra que resulta una generalización de la ecuación de Langevin, que rige el proceso de Uhlenbeck-Ornstein. A partir de la expresión analítica de las soluciones de ésta y de la original estudiamos las características estadísticas de ambos procesos solución, en particular los momentos de las distribuciones finito dimensionales, sus funciones de densidad de transición, las distribuciones límite y las condiciones de estacionariedad, obteniendo que la expresada generalización del proceso de U-O es el único proceso Gaussiano, Markoviano y estacionario no centrado en tiempo continuo. Por otra parte, se establece que las potencias del proceso lognormal-Gompertz generalizado satisfacen una E.D.E. del mismo tipo.

Palabras clave: Ecuación diferencial estocástica, ecuaciones de Kolmogorov, proceso log-normal, proceso de Gompertz, proceso de Uhlenbeck-Ornstein, ecuación de Langevin.



QÜESTIÓ, vol. 22, 1, pp. 39-68, 1998

MA

Un procedimiento para obtener clusters utilizando la D.V.S. de una matriz. Comparaciones con el biplot y con el modelo Q-factorial***A procedure of clustering using the SVD of a matrix. Comparisons with the biplot and with the Q-factor model***

J. L. González Caballero* y M. J. Valderrama Bonnet**

* Dpto. de Matemáticas. Facultad de Medicina. Universidad de Cádiz.

** Dpto. de Estadística e I.O. Facultad de Farmacia. Universidad de Granada.

Durante las últimas décadas, el análisis de un conjunto de n individuos medidos en p variables, proporcionando una matriz de datos $X_{n;p}$, mediante técnicas de representación que utilizan la Descomposición en Valores Singulares (DVS) de la matriz $X_{n;p}$ (o alguna derivada), han permitido resumir la información que aportan los datos en alguna forma óptima, siendo muy útil para indicar la presencia de clusters entre los n individuos y/o para prevenir ante posibles clasificaciones erróneas producidas por técnicas de agrupamiento más complejas. En este artículo estudiaremos un procedimiento que puede utilizarse en ocasiones para obtener clasificaciones naturales de un conjunto de datos, basado en la representación biplot y en el modelo Q-factorial que puede obtenerse a partir de la DVS.

Palabras clave: Descomposición en valores singulares, modelos factoriales, representación biplot, varianza generalizada, procedimientos de cluster no jerárquicos.



QÜESTIÓ, vol. 22, 1, pp. 145-156, 1998

OS

The post randomisation method for protecting microdata

J. Gouweleeuw, P. Kooiman, L. Willenborg and P. de Wolf

Statistics Netherlands. Division Research and Development. Department of Statistical Methods. The Netherlands.

This paper describes the Post Randomisation Method (PRAM) as a method for disclosure protection of microdata. Applying PRAM means that for each record in the data file according to a specified probability mechanism the score on a number of variables is changed. Since this probability mechanism is known, the characteristics of the latent true data can unbiasedly be estimated from the observed data moments in the perturbed file. Here PRAM is applied to categorical variables. It is shown that both crosstabulation and standard multivariate analysis techniques can easily be adapted to account for PRAM. It only requires pre-multiplication by the transpose of the inverted Markov transition matrix, specifying the randomisation process. Also, estimates for the additional variance introduced by PRAM are given. By a proper choice of the transition probabilities, PRAM can be applied in such a way that certain chosen marginal distributions in the original data file are left invariant in expectation. In that case the perturbed data can be used as if it were the original data. We describe how to obtain such an invariant PRAM process. Finally, some consequences of using PRAM in practice are discussed. The present paper is a shortened version of Kooiman *et al.* (1997).

Keywords: Post Randomisation Method (PRAM), disclosure, randomised response, Markov matrix, invariant matrix, perturbed data, data swapping.

○○○

QÜESTIÓ, vol. 26, 3, pp. 503-513, 2002

MA

La distància de l'Eixample

The Eixample Distance

M. Greenacre
Universitat Pompeu Fabra.

Inspirant-nos en el disseny de l'Eixample de Barcelona per Ildelfons Cerdà, definim una nova distància estadística, anomenada la distància de l'eixample. Estudiarem algunes propietats d'aquesta funció de distància i proposem algunes possibles aplicacions.

Paraules clau: Statistical distance, Manhattan distance, Euclidean distance.

○○○

QÜESTIÓ, vol. 25, 3, pp. 553-580, 2001

OS

El cens de població: un assaig d'interpretació mitjançant Data Mining

Population Census: an interpretation using Data Mining

C. Guisande Allende* i F. Subirada Curco**

* Institut d'Estadística de Catalunya.

** CEPBA-IBM Research Institute.

En aquest article es presenten els resultats d'un treball fet conjuntament entre l'Institut d'Estadística de Catalunya (Idescat), el Centre de Supercomputació de Catalunya (Cesca) i IBM. Les tres institucions van realitzar una anàlisi de les dades del Cens de població i habitatge de 1991 segons la metodologia Data Mining. Del conjunt de tècniques que es poden aplicar amb l'Intelligent Miner, es van fer servir les tècniques de segmentació i d'associació que s'han aplicat als individus segons el tipus de llar al que pertanyen.

Paraules clau: Data Mining, estadística oficial, associacions, segmentació, anàlisi de dades, aplicació, mètodes estadístics.

○○○

QÜESTIÓ, vol. 23, 3, pp. 425-436, 1999

RE

The relative potency of two preparations applied in the supplemented block design with multivariate responses

Z. Hanusz

Institute of Applied Mathematics. Agricultural University. Lublin. Poland.

The method of point and interval estimation of the relative potency of two preparations: Standard and Test in the multivariate case is presented. General formulae for testing the hypotheses about the parallelism of regression lines and the relative potency have been adopted to experiments in which doses of preparations are applied in the supplemented block designs. The designs of this kind, with two groups of treatments, the first group comprising the doses of the Standard preparation and the second group- the doses of the Test preparation have been used for bioassay. The details of the experimental plans as well as the test functions are presented. The theoretical considerations are illustrated with an example involving a generated data set.

Keywords: Relative potency, parallel-line design, supplemented block design, multivariate response.



QÜESTIIÓ, vol. 25, 3, pp. 509-520, 2001

RE

Practical Data Mining in a large utility company

G. Hébrail

Electricité de France, R&D Division.

We present in this paper the main applications of data mining techniques at Electricité de France, the French national electric power company. This includes electric load curve analysis and prediction of customer characteristics. Closely related with data mining techniques are data warehouse management problems: we show that statistical methods can be used to help to manage data consistency and to provide accurate reports even when missing data are present.

Keywords: Data mining, statistical methods, data warehouses, applications, load curve analysis, utility company.



QÜESTIIÖ, vol. 22, 2, pp. 231-242, 1998

BS

On maximum entropy priors and a most likely likelihood in auditing

A. Hernández Bastida*, M. C. Martel Escobar** and F. J. Vázquez Polo**

* Universidad de Granada.

** Universidad de Las Palmas de Gran Canaria.

There are two basic questions auditors and accountants must consider when developing testing and estimation applications using Bayes' Theorem: What prior probability function should be used and what likelihood function should be used. In this paper we propose to use

a maximum entropy prior probability function MEP with the most likely likelihood function MLL in the Quasi-Bayesian QB model introduced by McCray (1984). It is defined on a adequate parameter. Thus procedure only needs an expected value of q_0 known (in this paper the expected tainting) to obtain a MEP all an auditor or accountant need to supply are the range, as with any other prior, and the expected tainting, q_0 . We will see some practical applications of the methodology proposed about internal control evaluation in auditing.

Keywords: Maximum Entropy Prior, Partial Prior Information, Auditing.

○○○

SORT, vol. 31, 1, pp. 55-74, 2007

SE

Parameter estimation of S-distributions with alternating regresion

I-Chun Chou*, H. Martens** and E. O. Voit*

* Georgia Institute of Technology and Emory University.

** CIGENE/IKBM, Norwegian U. of Life Sciences.

We propose a novel 3-way alternating regression (3-AR) method as an effective strategy for the estimation of parameter values in S-distributions from frequency data. The 3-AR algorithm is very fast and performs well for error-free distributions and artificial noisy data obtained as randomsamples generated from S-distributions, as well as for traditional statistical distributions and for actual observation data. In rare cases where the algorithm does not immediately converge, its enormous speed renders it feasible to select several initial guesses and search settings as an effective countermeasure.

Keywords: Alternating regression, Parameter estimation, S-distribution, S-system.

○○○

SORT, vol. 29, 1, pp. 1-10, 2005

GS

Graphical display in outlier diagnostics; adequacy and robustness

N. K. Jajo

University of Western Sydney.

Outlier robust diagnostics (graphically) using Robustly Studentized Robust Residuals (RSRR) and Partial Robustly Studentized Robust Residuals (PRSRR) are established. One problem with some robust residual plots is that the residuals retain information from certain predicated values (Velilla, 1998). The RSRR and PRSRR techniques are unaffected by this complication and as a result they provide more interpretable results.

Keywords: Masking, outlier, robust diagnostics, robust residuals, swamping.

○○○

QÜESTIÓ, vol. 24, 1, pp. 137-150, 2000

OS

Comparative analysis of alternative sampling plans to create a farm accountancy data network for the agricultural sector of Navarra

L. Júdez and C. Chaya

Unidad de Estadística. Departamento de Economía y Ciencias Sociales Agrarias. Escuela Técnica Superior de Ingenieros Agrónomos de Madrid.

This study presents the method that was followed and the results of the analysis for establishing a sampling plan for the Farm Accountancy Data Network (FADN) of the commercial agricultural sector of the Spanish Autonomous Community of Navarra. The first part of the study presents the categories considered of the different stratification criteria for the commercial farms of Navarra: geographical units (subregions), types of farming (TF) and economic size in ESU (European Size Unit). Then the authors define sampling plans with different objectives to be compared. These objectives are: i) maximum accuracy in the estimation of the Standard Gross Margin (SGM) for the whole of the commercial agricultural sector, ii) the same accuracy in the various types of farming (TF) and iii) the same accuracy in the individual strata. Special attention is given to the effects of introducing geographical units as a stratification criterion. Given the sample size and the characteristics of the population of commercial farms in Navarra the plan without geographical stratification that gives approximately the same accuracy in each TF seems to be the most suitable sampling plan for the FADN of this region. Even though this study is limited to Navarra, it may be of help when considering sampling plans for FADN not only in other Autonomous Communities but also on a national scale.

Keywords: Farm accountancy data network (FADN), stratified sampling, sample surveys, Navarra.

○○○

SORT, vol. 28, 1, pp. 1-8, 2004

SE

On-line nonparametric estimation

R. Khasminskii

Department of Mathematics. Wayne State University. USA.

Survey of some recent results on nonparametric on-line estimation is presented. The first result deals with an on-line estimation for a smooth signal $S(t)$ in the classic 'signal plus Gaussian White Noise (GWN)' model. Then an analogous on-line estimator for the regression estimation problem with equidistant design is described and justified. Finally some preliminary results related to the on-line estimation for the diffusion observed process are described.

Keywords: White Gaussian Noise, on-line estimation, nonlinear filter.

○○○

SORT, vol. 28, 2, pp. 201-214, 2004

PD

Some discrete exponential dispersion models: Poisson-Tweedie and Hinde-Demétrio classes

C. Kokonendji*, S. Dossou-Gbété* and C. Demétrio**

* University of Pau-LMA, Pau, France.

** University of Sao Paulo. Brazil.

In this paper we investigate two classes of exponential dispersion models (EDMs) for overdispersed count data with respect to the Poisson distribution. The first is a class of Poisson mixture with positive Tweedie mixing distributions. As an approximation (in terms of unit variance function) of the first, the second is a new class of EDMs characterized by their unit variance functions of the form $\mu + \mu p$, where p is a real index related to a precise model. These two classes provide some alternatives to the negative binomial distribution ($p = 2$) which is classically used in the framework of regression models for count data when overdispersion results in a lack of fit of the Poisson regression model. Some properties are then studied and the practical usefulness is also discussed.

Keywords: Negative binomial distribution, overdispersion, Poisson mixture, Tweedie family, unit variance function.

○○○

SORT, vol. 28, 2, pp.111-124, 2004

PD

On invariant density estimation for ergodic diffusion processes

Y. A. Kutoyants

Université du Maine, Le Mans.

We present a review of several results concerning invariant density estimation by observations of ergodic diffusion process and some related problems. In every problem we propose a lower mínimas bound on the risks of all estimators and then we construct an asymptotically efficient estimator.

Keywords: Ergodic diffusion, density estimation, large sample theory, efficient estimation.

○○○

QÜESTIÓ, vol. 26, 1-2, pp. 175-196, 2002

OS

Alimentación de modelos cuantitativos con información subjetiva: aplicación Delphi en la elaboración de un modelo de imputación del gasto turístico individual en Catalunya

Supplying of subjective information to quantitative models: the Delphi Method applied to a model of attribution of the individual tourist expense in Catalunya

J. Landeta, J. Matey, V. Ruiz y O. Villarreal

Universidad del País Vasco/Euskal Herriko Unibertsitatea. Instituto de Economía Aplicada a la Empresa. Facultad de Ciencias Económicas y Empresariales.

El presente artículo presenta un estudio realizado para el Institut d'Estadística de Catalunya, donde se ha aplicado una técnica que trabaja con información subjetiva (el Método Delphi) para obtener unos datos que puedan ser empleados en la alimentación parcial de un modelo matemático, que se nutre principalmente de datos estadísticos (objetivos), con el fin de incrementar la utilidad global del modelo. El artículo justifica la utilización en determinadas circunstancias de la información subjetiva, describe el método Delphi y plantea una serie de proposiciones referentes a la utilidad, empleo y mejora de esta técnica que se ven refrendadas en el caso expuesto.

Palabras clave: Método Delphi, información subjetiva, tecnologías de la información.

○○○

SORT, vol. 29, 1, pp. 11-26, 2005

PD

Positivity theorem for a general manifold

R. Léandre

Institut de Mathématiques. Faculté des Sciences. Université de Bourgogne.

We give a generalization in the non-compact case to various positivity theorems obtained by Malliavin Calculus in the compact case.

Keywords: Malliavin calculus, density.

○○○

QÜESTIÓ, vol. 23, 2, pp. 321-342, 1999

OS

Qualité de l'information dans les enquêtes

Quality of data in sample surveys

L. Lebart

Centre National de la Recherche Scientifique. Ecole Nationale Supérieure des Télécommunications. Paris.

Cet article tente de montrer les contributions des analyses statistiques multidimensionnelles et des analyses textuelles à l'amélioration de la qualité des résultats d'une enquête. L'idée de base est la suivante: chaque phase de l'enquête et de sa réalisation sur le terrain peut donner à la mesure ou au relevé de certaines variables de contrôles. Le fichier des données individuelles se voit alors doublé d'un fichier de variables de contrôles, pouvant être numériques (chronométrage de l'entrevue) nominales (circonstances de l'entrevue, appréciations) ou textuelles (commentaires libres de l'enquêteur et de l'enquêté). Les méthodes précitées permettent alors de visualiser et de confronter ces deux fichiers, et donc de

juger la cohérence globale de l'information, en positionnant les variables de contrôles parmi les variables de l'enquête.

Mots clés: Qualité de l'information, Enquêtes socio-économiques, Analyse des dones, Analyses statistiques textuelles.



SORT, vol. 28, 2, pp. 177-190, 2004

TS

Asymptotically optimal filtering in linear systems with fractional Brownian noises

A. Le Breton, M. L. Kleptsyna and M. Viot

Laboratoire de Modélisation et Calcul/Université J. Fourier. Grenoble. France.

In this paper, the filtering problem is revisited in the basic Gaussian homogeneous linear system driven by fractional Brownian motions. We exhibit a simple approximate filter which is asymptotically optimal in the sense that, when the observation time tends to infinity, the variance of the corresponding filtering error converges to the same limit as for the exact optimal filter.

Keywords: Fractional Brownian motion, homogeneous linear system, optimal filtering, filtering error, asymptotic variance.



SORT, vol. 29, 2, pp. 141-182, 2005

SI

Likelihood for random-effect models

Y. Lee* and J. A. Nelder**

* Seoul National University.

** Imperial College London, U.K.

For inferences from random-effect models Lee and Nelder (1996) proposed to use hierarchical likelihood (h-likelihood). It allows inference from models that may include both fixed and random parameters. Because of the presence of unobserved random variables h-likelihood is not a likelihood in the Fisherian sense. The Fisher likelihood framework has advantages such as generality of application, statistical and computational efficiency. We introduce an extended likelihood framework and discuss why it is a proper extension, maintaining the advantages of the original likelihood framework. The new framework allows likelihood inferences to be drawn for a much wider class of models.

Keywords: Generalized linear models, hierarchical models, h-likelihood.



QÜESTIIÖ, vol. 22, 2, pp. 313-335, 1998

NO

Un método primal de optimización semi-infinita para la aproximación uniforme de funciones***A primal semi-infinite programming method for the uniform approximation problem***

T. León, S. Sanmatías y E. Vercher

Departament d'Estadística i Investigació Operativa. Facultat de Matemàtiques. Universitat de València.

En este trabajo presentamos un algoritmo que resuelve problemas clásicos de aproximación que pueden ser formulados como programas semi-infinitos lineales. Hemos estudiado la caracterización algebraica de los puntos extremos y demostrado algunas de sus propiedades. Hemos diseñado un procedimiento que genera direcciones factibles a partir de la solución de ciertos programas lineales finitos, que también caracteriza la solución óptima del problema. El método incorpora una etapa interna de purificación para alcanzar un punto extremo desde cualquier solución factible, mejorando el valor de la función objetivo. Finalmente, comparamos el comportamiento de las diferentes estrategias sobre varios problemas de aproximación, comprobando la eficacia del algoritmo propuesto.

Palabras clave: Programación semi-infinita, aproximación uniforme, puntos extremos, métodos de direcciones factibles.

○○○

QÜESTIIÖ, vol. 24, 1, pp. 111-136, 2000

OS

Disseny i construcció d'una mostra estratificada a partir de dades censals***Design and construction of a stratified sample from census data***

P. López, C. Lozares i M. Domínguez

Grup d'Estudis Sociològics sobre la Vida Quotidiana i el Treball (QUIT). Departament de Sociologia. Universitat Autònoma de Barcelona.

L'objectiu de l'article és presentar les principals característiques del disseny i del procés de construcció de la mostra estratificada de l'"Enquesta metropolitana de Barcelona. Condicions de vida i hàbits de la població" de l'edició de l'any 1995. A partir de la informació que prové del cens de població s'agrupen a les persones en seccions censals i es procedeix a la construcció dels estrats d'acord amb un procediment on s'impliquen tècniques d'anàlisi multivariable: d'anàlisi factorial de components principals i de classificació automàtica. Finalment, s'afixa una grandària de mostra segons el criteri òptim de Neyman. El procediment seguit i els criteris i decisions que s'utilitzen van més enllà dels resultats habituals esperats a tota mostra, doncs proporciona conclusions d'interès per a l'anàlisi sociològica i la planificació del territori, al mateix temps que serveix de criteri de validació de les conclusions posteriors a l'anàlisi de l'enquesta.

Paraules clau: Enquesta per mostreig, anàlisi de dades, anàlisi de components principals, anàlisi de classificació.

○○○

QÜESTIÓ, vol. 24, 3, pp. 495-528, 2000

NF

Del poliedro del agente viajero gráfico al de rutas de vehículos con demanda compartida***From the graphical traveling salesman polyhedron to the one associated with the split delivery vehicle routing problem***

C. Martínez y E. Mota

Departament d'Estadística i Investigació Operativa. Universitat de València.

En este trabajo abordamos el estudio del poliedro asociado al Problema de Rutas de Vehículos con Demanda Compartida, problema de distribución que surge cuando hay que repartir mercancías a un conjunto de clientes utilizando una flota fija de vehículos de capacidad limitada. El objetivo es diseñar las rutas de forma que se minimice la distancia total recorrida. Se diferencia de otros problemas más conocidos de rutas con capacidades en que se permite abastecer la demanda de cada cliente utilizando más de un vehículo. Presentamos resultados poliédricos que se obtienen como extensión de otros previamente conocidos para el Problema del Agente Viajero Gráfico.

Palabras clave: Rutas de vehículos, demanda compartida, poliedro, facetas.



QÜESTIÓ, vol. 23, 2, pp. 343-362, 1999

OS

Diseño de encuestas de opinión: barómetro CIS***Surveys Design for Opinions: Barometer CIS***

V. Martínez

Departamento de Banco de Datos. Centro de Investigaciones Sociológicas (CIS). Madrid.

El diseño de las encuestas de opinión puede incluirse dentro de los diseños generales de las encuestas sobre población, cuyo objetivo es la obtención de información mediante la observación de los individuos que la componen. Las características diferenciales con respecto a los diseños generales son la utilización de tamaños de muestra reducidos y una selección no probabilística de los individuos. En este sentido, la calidad de los resultados obtenidos suele ser cuestionada y, por tanto, las extrapolaciones sobre la población deben ser tomadas con cautela. En este artículo se revisan y valoran estos preceptos en el marco de la encuesta de opinión que elabora mensualmente el Centro de Investigaciones Sociológicas-CIS (conocida como Barómetro CIS), cuya versión inicial fue presentada en las Primeras Jornadas Internacionales "Generación de Información Estadística: Calidad y Limitaciones" (Barcelona, 30 noviembre-1 diciembre 1998) que organizó la Xarxa Temàtica "Enquestes i Qualitat de la Informació Estadística".

Palabras clave: Marcos, cuestionario, diseño, tamaño muestral, unidades de muestreo, conglomerados, estratificación, afijación, tamaño del conglomerado, cuotas, estimadores, costes, calidad.



SORT, vol. 28, 2, pp. 161-176, 2004

TS

Extremes of periodic moving averages of random variables with regularly varying tail probabilities

A. P. Martins and H. Ferreira

Departamento de Matemática (Centro de Matemática), Universidade da Beira Interior. Portugal.

We define a family of local mixing conditions that enable the computation of the extremal index of periodic sequences from the joint distributions of k consecutive variables of the sequence. By applying results, under local and global mixing conditions, to the $(2m - 1)$ -dependent periodic sequence $X_n^{(m)} = \sum_{j=-m}^{m-1} c_j Z_{n-j}$, $n \geq 1$, we compute the extremal index of the periodic moving average sequence $X_n = \sum_{j=-\infty}^{\infty} c_j Z_{n-j}$, $n \geq 1$, of random variables with regularly varying tail probabilities.

This paper generalizes the theory for extremes of stationary moving averages with regularly varying tail probabilities.

Keywords: Periodic moving average processes, extremal index, mixing condition.

○○○

QÜESTIÓ, vol. 22, 3, pp. 511-526, 1998

OS

A comparative study of microaggregation methods

J. M. Mateo-Sanz, and J. Domingo-Ferrer

Departament d'Enginyeria Informàtica. Universitat Rovira i Virgili.

Microaggregation is a statistical disclosure control technique for microdata. Raw microdata (i. e. individual records) are grouped into small aggregates prior to publication. Each aggregate should contain at least k records to prevent disclosure of individual information. Fixed-size microaggregation consists of taking fixed-size microaggregates (size k). Data-oriented microaggregation (with variable group size) was introduced recently. Regardless of the group size, microaggregates on a multidimensional data set can be formed using univariate techniques on projected data or using multivariate techniques. This paper presents the first method for multivariate fixed-size microaggregation. In addition, a real data set is used to compare the information loss and output data quality of fixed-size vs. data-oriented, and univariate vs. multivariate microaggregation.

Keywords: Statistical disclosure control, Microaggregation, Data-oriented, microaggregation, Microdata protection.

○○○

QÜESTIÓ, vol. 26, 1-2, pp. 29-59, 2002

GS

Statistical procedures for spatial point pattern recognition

J. Mateu

Department of Mathematics, Universitat Jaume I.

Spatial structures in the form of point patterns arise in many different contexts, and in most of them the key goal concerns the detection and recognition of the underlying spatial pattern. Particularly interesting is the case of pattern analysis with replicated data in two or more experimental groups. This paper compares design-based and model-based approaches to the analysis of this kind of spatial data. Basic questions about pattern detection concern estimating the properties of the underlying spatial point process within each experimental group, and comparing the properties between groups. The paper discusses how either approach can be implemented in the specific context of a single-factor replicated experiment and uses simulations to show how the model-based approach can be more efficient when the underlying model assumptions hold, but potentially misleading otherwise.

Keywords: Expected significance levels, Replicated patterns, Spatial point structures.



QÜESTIÓ, vol. 22, 1, pp. 175-194, 1998

BI

Modelización de la distribución espacial de quistes en el estómago de la marsopa mediante un proceso de Gibbs***Modelling the spatial distribution of cysts placed in a porpoise stomach using a Gibbs process***

J. Mateu* y F. Montes**

* Departamento de Matemáticas de la Universitat Jaume I.

** Departamento de Estadística e I.O. de la Universitat de Valencia.

En un trabajo previo (Balaguer *et al.*, 1992) los autores estudiaron la distribución espacial de los quistes en el estómago de la marsopa para contrastar su compatibilidad con la hipótesis de aleatoriedad espacial completa (AEC). El resultado condujo a rechazar la hipótesis señalando en la dirección de un patrón agregado, confirmando la sospecha inicial de los zoólogos. Este trabajo completa aquel estudio en la dirección que parece más coherente a sus autores: la de conjeturar un modelo puntual espacial con interacciones dos a dos (un proceso de Strauss) y estimar sus parámetros. Previamente se lleva a cabo el imprescindible desarrollo teórico poniendo el énfasis en los distintos métodos de estimación máximo-verosímil.

Palabras clave: Procesos puntuales finitos, procesos de Gibbs, procesos de Strauss.



QÜESTIÓ, vol. 24, 1, pp. 3-25, 2000

SE

Approximate maximum likelihood estimation for a spatial point pattern

J. Mateu* and F. Montes**

* Departament de Matemàtiques. Universitat Jaume I. Castelló.

** Departament de Estadística i I.O. Universitat de València.

Several authors have proposed stochastic and non-stochastic approximations to the maximum likelihood estimate for a spatial point pattern. This approximation is necessary because of the difficulty of evaluating the normalizing constant. However, it appears to be neither a general theory which provides grounds for preferring a particular method, nor any extensive empirical comparisons. In this paper, we review five general methods based on approximations to the maximum likelihood estimate which have been proposed in the literature. We also present the results of a comparative simulation study developed for the Strauss model.

Keywords: Gibbs distribution, maximum likelihood, Monte Carlo inference, stochastic approximation, Strauss model.



QÜESTIÓ, vol. 26, 1-2, pp. 61-85, 2002

SE

Detección de rasgos en imágenes binarias mediante procesos puntuales espaciales marcados

Detection of features in binary images by means of spatial marked point processes

J. Mateu y G. Lorenzo

Departamento de Matemáticas, Universitat Jaume I.

En este trabajo consideramos el problema de la detección de rasgos bajo la presencia de ruido en imágenes que tras un cierto tratamiento se reducen a binarias, por la presencia de dos tipos de elementos. Podemos encontrar ejemplos de este problema en la detección de minas por medio de imágenes de avión o satélite, en la búsqueda de rasgos en imágenes microscópicas de células, o en la caracterización de fallas en zonas de terremotos. En primer lugar revisamos algunos métodos de detección jerárquicos basados en modelos probabilísticos, en los que los rasgos proceden de distribuciones normales multivariantes y el ruido surge según un proceso de Poisson espacial. Posteriormente, presentamos una nueva solución al problema mediante el uso de procesos puntuales espaciales marcados. Definimos un proceso puntual marcado en el que a cada localización se le asigna un par de marcas: las distancias al K -ésimo vecino más cercano y una variable dicotómica diferenciadora del rasgo frente al ruido. Estas distancias se modelizan como una mixtura de distribuciones cuyos parámetros se determinan mediante el algoritmo EM. Finalmente, la nueva metodología es evaluada y contrastada sobre simulaciones y casos reales.

Palabras clave: Algoritmo EM, mixturas de distribuciones, proceso de Poisson, procesos puntuales espaciales.



QÜESTIÓ, vol. 26, 1-2, pp. 87-108, 2002

ST

Estimación de la función de distribución sobre poblaciones finitas mediante diseños muestrales bietápicos apropiados***Estimating distributions functions using appropriate sampling designs in two stage cluster sampling***

J. A. Mayor y M. Martínez

Dpto. de Estadística e Investigación Operativa. Universidad de Sevilla.

Con el objeto de estimar la función de distribución de una variable de estudio sobre una población finita, se propone en este trabajo emplear el estimador de Horvitz-Thompson, lo que proporciona una estrategia muestral insesgada, siendo la varianza de dicho estimador una función real de variable real cuya minimización permite obtener diseños muestrales óptimos bajo diferentes criterios. En este trabajo empleamos la norma $\|\cdot\|_1$ como criterio de optimización, minimizando la norma de la varianza, como función de la matriz del diseño muestral. De esta forma, suponiendo muestreo por conglomerados en dos etapas y considerando como dominio de búsqueda el conjunto de los diseños muestrales de tipo uniforme, en el sentido de ser iguales las probabilidades de inclusión de primer orden, se estudia la obtención de diseños muestrales adecuados para dicha estimación.

Palabras clave: Muestreo, poblaciones finitas, diseño muestral, función de distribución, conglomerados.

○○○

SORT, vol. 27, 1, pp. 95-112, 2003

NF

Optimization of touristic distribution networks using genetic algorithms

J. R. Medina* and V. Yepes**

* Universidad Politécnica de Valencia.

** Agència Valenciana del Turisme.

The eight basic elements to design genetic algorithms (GA) are described and applied to solve a low demand distribution problem of passengers for a hub airport in Alicante and 30 touristic destinations in Northern Africa and Western Europe. The flexibility of GA and the possibility of creating mutually beneficial feed-back processes with human intelligence to solve complex problems as well as the difficulties in detecting erroneous codes embedded in the software are described. A new three-parent edge mapped recombination operator is used to solve the capacitated vehicle routing problem required for estimating associated costs with touristic distribution networks of low demand. GA proved to be very flexible especially in changing business environments and to solve decision making problems involving ambiguous and sometimes contradictory constraints.

Keywords: Distribution networks, vehicle routing problem, tourism demand, air transportation, genetic algorithms, edge mapped recombination operator.

○○○

QÜESTIÓ, vol. 24, 2, pp. 335-365, 2000

OS

L'Estadística Oficial i l'impacte de l'economia digital emergent

Official statistics and the impact of the emergent digital economy

V. Meléndez

Institut d'Estadística de Catalunya.

La implantació de les tecnologies de la informació i les comunicacions (TIC) en totes les esferes socials, fa necessària la mesura del seu impacte econòmic i social i del grau d'assoliment de la societat de la informació. Es presenta una anàlisi de la bibliografia publicada en relació amb la incidència sobre la metodologia estadística que se'n deriva, així com de les ampliacions i millores que ja s'estan duent a terme en aquest camp en els països capdavanters. Es fa esment específic a la qüestió de l'actualització de les classificacions d'activitats econòmiques en aquesta etapa. Finalment, es descriuen els intents de constituir conjunts d'indicadors que donin compte i ajudin a comprendre aquests fenòmens, en especial els duts a terme a l'Idescat.

Paraules clau: Tecnologies de la informació i les comunicacions (TIC), economia digital, classificacions estadístiques, societat de la informació, indicadors econòmics i mesures.



QÜESTIÓ, vol. 23, 1, pp. 61-83, 1999

NO

On the compatibility of classical multiplier estimates with variable reduction techniques when there are nonlinear inequality constraints

E. Mijangos* and N. Nabona**

* Department of Applied Mathematic, Statistics and Operations Research, Zientzi Fakultatea (EHU), P.O. Bilbao.

** Department of Statistics and Operations Research, Universitat Politècnica de Catalunya.

The minimization of a nonlinear function subject to linear and nonlinear equality constraints and simple bounds can be performed through minimizing a partial augmented Lagrangian function subject only to linear constraints and simple bounds by variable reduction techniques. The first-order procedure for estimating the multiplier of the nonlinear equality constraints through the Kuhn-Tucker conditions is analyzed and compared to that of Hestenes-Powell. There is a method which identifies those major iterations where the procedure based on the Kuhn-Tucker conditions can be safely used and also computes these estimates. This work justifies the extension of the former results to the case of general inequality constraints. To this end two procedures that convert inequalities into equalities are considered.

Keywords: Nonlinear programming, general inequality constraints, augmented lagrangian, variable reduction, Lagrangemultiplier estimates.



QÜESTIÓ, vol. 26, 1-2, pp. 141-171, 2002

NO

On superlinear multiplier update methods for partial augmented Lagrangian techniques

E. Mijangos

Department of Applied Mathematics and Statistics and O.R., Zientzi Fakultatea (UPV/EHU).

The minimization of a nonlinear function with linear and nonlinear constraints and simple bounds can be performed by minimizing an augmented Lagrangian function, including only the nonlinear constraints. This procedure is particularly interesting in case that the linear constraints are flow conservation equations, as there exist efficient techniques to solve nonlinear network problems. It is then necessary to estimate their multipliers, and variable reduction techniques can be used to carry out the successive minimizations. This work analyzes the possibility of estimating these multipliers using Newton-like methods. Several procedures are put forward which combine first and second-order estimation, and are compared with each other and with the Hestenes-Powell multiplier estimation by means of computational tests.

Keywords: Partial Augmented Lagrangian, Variable Reduction, Lagrange Multipliers, Superlinear-order Estimation, Nonlinear Network.



QÜESTIÓ, vol. 26, 1-2, pp. 259-271, 2002

BI

Inverse sampling and triangular sequential designs to compare a small proportion with a reference value

V. Moreno**, I. Martín***, F. Torres**, M. Horas**, J. Ríos** and J. R. González*

* Instituto Catalán de Oncología. Barcelona.

** Laboratorio de Bioestadística y Epidemiología. Facultad de Medicina, Universidad Autónoma de Barcelona.

*** Departamento de Estadística y Econometría. Universidad Carlos III. Madrid.

Inverse sampling and formal sequential designs may prove useful in reducing the sample size in studies where a small population proportion p is compared with a hypothesized reference proportion p_0 . These methods are applied to the design of a cytogenetic study about chromosomal abnormalities in men with a daughter affected by Turner's syndrome. First it is shown how the calculated sample size for a classical design depends on the parameterization used. Later this sample size is compared with the required sample size in an inverse sampling design and a triangular sequential design using four different parameterizations (absolute differences, log-odds ratio, angular transform and Sprott's transform). The expected savings in sample size, when the alternative hypothesis is true, are 20% of the fixed sample size for the inverse sampling design and 40% for the triangular sequential design.

Keywords: Sample size, inverse sampling, sequential methods, triangular sequential test.



SORT, 30, 1, pp. 3-54, 2006

BS

On the frequentist and Bayesian approaches to hypothesis testing

E. Moreno and F. J. Girón
University of Granada and University of Málaga.

Hypothesis testing is a model selection problem for which the solution proposed by the two main statistical streams of thought, frequentists and Bayesians, substantially differ. One may think that this fact might be due to the prior chosen in the Bayesian analysis and that a convenient prior selection may reconcile both approaches. However, the Bayesian robustness viewpoint has shown that, in general, this is not so and hence a profound disagreement between both approaches exists. In this paper we briefly revise the basic aspects of hypothesis testing for both the frequentist and Bayesian procedures and discuss the variable selection problem in normal linear regression for which the discrepancies are more apparent. Illustrations on simulated and real data are given.

Keywords: Bayes factor, consistency, intrinsic priors, loss function, model posterior probability, p -values.



QÜESTIIÓ, vol. 25, 2, pp. 263-284, 2001

MA

El sesgo condicionado en el análisis de influencia: una revisión***Conditional bias in influence analysis: a review***

J. M. Muñoz-Pichardo, J. L. Moreno-Rebollo, T. Gómez-Gómez y A. Enguix-González
Universidad de Sevilla. Facultad de Matemáticas. Departamento de Estadística e Investigación Operativa.

El sesgo condicionado se ha propuesto como diagnóstico de influencia en distintos modelos y técnicas estadísticas. Tratando de recoger una visión global de la utilidad del concepto, en este trabajo se hace una revisión general del mismo relacionándolo con la curva de sensibilidad y la curva de influencia muestral. Además, se señalan posibles líneas de trabajo que permitirán abordar el análisis de la influencia a través de este enfoque en una gran variedad de técnicas estadísticas.

Palabras clave: Análisis de influencia, modelos lineales, componentes principales, muestreo en poblaciones finitas, sesgo condicionado.



SORT, vol. 29, 1, pp. 43-56, 2005

ID

Information matrices for some elliptically symmetric distributions

S. Nadarajah* and S. Kotz**

* University of Nebraska.

** The George Washington University.

The Fisher information matrices are derived for three of the most popular elliptically symmetric distributions: the Pearson type II, Pearson type VII and the Kotz type distributions. We hope the results could be important to the many researchers working in this area.

Keywords: Elliptically symmetric Kotz type distribution, Elliptically symmetric Pearson type II distribution, Elliptically symmetric Pearson type VII distribution, Fisher information matrices.



SORT, vol. 29, 2, pp. 183-200, 2005

PD

Muliere and Scarsini's bivariate Pareto distribution: sums, products, and ratios

S. Nadarajah* and S. Kotz**

* University of Nebraska.

** The George Washington University.

We derive the exact distributions of $R = X + Y$, $P = X Y$ and $W = X/(X + Y)$ and the corresponding moment properties when X and Y follow Muliere and Scarsini's bivariate Pareto distribution. The expressions turn out to involve special functions. We also provide extensive tabulations of the percentage points associated with the distributions. These tables –obtained using intensive computing power– will be of use to practitioners of the bivariate Pareto distribution.

Keywords: Incomplete beta function, Gauss hypergeometric function, Muliere and Scarsini's bivariate Pareto distribution, products of random variables, ratios of random variables, sums of random variables.



QÜESTIÓ, vol. 25, 3, pp. 415-436, 2001

TS

Detección de M señales gaussianas utilizando el desarrollo modificado de un proceso estocástico***Detection of M gaussian signals using the modified approximate Karhunen-Loève expansion of a stochastic process***

J. Navarro-Moreno y J. C. Ruíz-Molina

Departamento de Estadística e I.O. Universidad de Jaén.

Utilizando el desarrollo modificado de un proceso estocástico se propone una nueva metodología, alternativa a la basada en el desarrollo de Karhunen-Loève, para el problema de detección de M señales Gaussianas en ruido Gaussiano blanco. Las soluciones proporcionadas no presentan el problema del cálculo de los autovalores y autofunciones asociados a la función de covarianza involucrada y son fácilmente implementables desde el punto de vista práctico.

Palabras clave: Desarrollo modificado de un proceso estocástico, detección de M señales gaussianas en ruido gaussiano blanco.

○○○

SORT, vol. 27, 1, pp. 41-64, 2003

BI

Survival analysis with coarsely observed covariates

S. F. Nielsen
University of Copenhagen.

In this paper we consider analysis of survival data with incomplete covariate information. We model the incomplete covariates as a random coarsening of the complete covariate, and an overview of the theory of coarsening at random is given. Various ways of estimating the parameters of the model for the survival data given the covariates are discussed and compared.

Keywords: Incomplete data, ignorability, efficiency, EM algorithm, martingale estimating function, inverse probability weighting.

○○○

QÜESTIIÓ, vol. 24, 2, pp. 243-249, 2000

GE

A note on the scalar Haffian

H. Neudecker
Cesaro, Schagen, Department of Economics, University of Amsterdam.

In this note a uniform transparent presentation of the scalar Haffian will be given. Some well-known results will be generalized. A link will be established between the scalar Haffian and the derivative matrix as developed by Magnus and Neudecker.

Keywords: Magnus-Neudecker derivative matrix, matrix vectorization, Kronecker product, duplication matrix, commutation matrix.

○○○

QÜESTIÓ, vol. 24, 3, pp. 419-424, 2000

GE

A note on the matrix Haffian

H. Neudecker

Cesaro, Schagen, Department of Economics, University of Amsterdam.

This note contains a transparent presentation of the matrix Haffian. A basic theorem links this matrix and the differential of the matrix function under investigation, viz $\nabla F(X)$ and $dF(X)$. Frequent use is being made of matrix derivatives as developed by Magnus and Neudecker.

Keywords: Matrix differentials, matrix functions.

○○○

QÜESTIÓ, vol. 25, 2, pp. 187-210, 2001

GE

Some applications of the matrix Haffian in connection with differentiable matrix functions of a central wishart variate

H. Neudecker

Cesaro, Schagen, Department of Economics, University of Amsterdam.

In this paper we revisit Haff 's seminal work on the matrix Haffian as we proposed to call it. We review some results, and give new derivations. Use is made of the link between the matrix Haffian ∇F and the differential of the matrix function, dF .

Keywords: Kronecker product, commutation matrix, Hadamard product, matrix differentiation, matrix differentials, matrix partitioning.

○○○

Sort, vol. 27, 2, pp. 153-164, 2003

GE

On two matrix derivatives by Kollo and von Rosen

H. Neudecker

Cesaro, Schagen, Department of Economics, University of Amsterdam.

The article establishes relationships between the matrix derivatives of F with respect to X as introduced by von Rosen (1988), Kollo and von Rosen (2000) and the Magnus-Neudecker (1999) matrix derivative. The usual transformations apply and the Moore-Penrose inverse of the duplication matrix is used. Both X and F have the same dimension.

Keywords: Derivades i diferencials de matrius, vectorització, commutació i duplicació de matrius.

○○○

SORT, vol. 28, 1, pp. 27-36, 2004

MA

On best affine unbiased covariance-preserving prediction of factor scores

H. Neudecker

Cesaro, Schagen, Department of Economics, University of Amsterdam.

This paper gives a generalization of results presented by ten Berge, Krijnen, Wansbeek & Shapiro. They examined procedures and results as proposed by Anderson & Rubin, McDonald, Green and Krijnen, Wansbeek & ten Berge. We shall consider the same matter, under weaker rank assumptions. We allow some moments, namely the variance Ω of the observable scores vector and that of the unique factors, Ψ , to be singular. We require $T' \Psi T > 0$, where $T \Lambda T'$ is a Schur decomposition of Ω . As usual the variance of the common factors, Φ , and the loadings matrix A will have full column rank.

Keywords: Factor analysis, factor scores, covariance-preserving, Kristof-type theorem.

○○○

SORT, vol. 28, 2, pp. 191-200, 2004

GE

Estimation of the noncentrality matrix of a noncentral Wishart distribution with unit scale matrix. A matrix generalization of Leung's domination result

H. Neudecker

Cesaro, Schagen, Department of Economics, University of Amsterdam.

The main aim is to estimate the noncentrality matrix of a noncentral Wishart distribution. The method used is Leung's but generalized to a matrix loss function. Parallely Leung's scalar noncentral Wishart identity is generalized to become a matrix identity. The concept of Löwner partial ordering of symmetric matrices is used.

Keywords: Noncentral Wishart matrix identity, noncentrality matrix, decision-theoretic estimation, matrix loss function, Löwner matrix ordering, Haffians.

○○○

SORT, vol. 30, 1, pp. 85-90, 2006

GE

A matrix function useful in the estimation of linear continuous-time models

H. Neudecker

Cesaro, Schagen, Department of Economics, University of Amsterdam.

In a recent publication Chen & Zadrozny (2001) derive some equations for efficiently computing e^A and ∇e^A , its derivative. They employ an expression due to Bellman (1960), Snider (1964) and Wilcox (1967) for the differential de^A and a method due to Van Loan (1978) to find the derivative ∇e^A . The present note gives a) a short derivation of ∇e^A by way of the

Bellman-Snider-Wilcox result, *b*) a shorter derivation without using it. In both approaches there is no need for Van Loan's method.

Keywords: Matrix derivatives, vectorization, matrix exponential.

○○○

SORT, vol. 31, 2, pp. 201-206, 2007

PD

A recursion formula for expected negative and positive powers of the central Wishart distribution

H. Neudecker

Cesaro, Schagen, Department of Economics, University of Amsterdam.

We use Haff's fundamental identity to express the expectation of S^p in lower-order terms, where S follows the central Wishart distribution.

Keywords: Haff's Fundamental Identity, matrix differentiation, central Wishart distribution.

○○○

SORT, vol. 30, 1, pp. 91-100, 2006

SE

About one problem of Bernoulli and Euler from the theory of statistical estimation

M. Nikulin

Université Bordeaux.

We consider some results by D. Bernoulli and L. Euler on the method of maximum likelihood in parametric estimation. The statistical analysis is made by considering a parametric family with a shift parameter.

Keywords: Asymptotic normality, sample mean, Bernoulli's estimator, Euler's estimator, maximum likelihood, unbiased estimator.

○○○

QÜESTIÓ, vol. 25, 2, pp. 225-262, 2001

TS

Modelización de datos longitudinales con estructura de covarianza no estacionarias: modelos de coeficientes aleatorios frente a modelos alternativos

Modelling longitudinal data with nonstationary covariance structures: random coefficients models versus alternative models

V. Núñez-Antón* y D. L. Zimmerman**

* Departamento de Econometría y Estadística (E.A. III). Universidad del País Vasco.

** Department of Statistics and Actuarial Science. The University of Iowa.

Un tema que ha suscitado el interés de los investigadores en datos longitudinales durante las dos últimas décadas, ha sido el desarrollo y uso de modelos paramétricos explícitos para la estructura de covarianza de los datos. Sin embargo, el análisis de estructuras de covarianza no estacionarias en el contexto de datos longitudinales no se ha realizado de forma detallada principalmente debido a que las distintas aplicaciones no hacían necesario su uso. Muchos son los modelos propuestos recientemente, pero la mayoría son estacionarios de segundo orden. Algunos de éstos, sin embargo, son no estacionarios y suficientemente flexibles, de tal forma que es posible modelizar varianzas no constantes y/o correlaciones que no sean sólo función del tiempo que separa a dos observaciones dadas. Estudiaremos algunas de estas propuestas y las compararemos con los modelos de coeficientes aleatorios, evaluando sus ventajas y desventajas e indicando cuándo su uso no es apropiado o útil. Presentaremos dos ejemplos para ilustrar el ajuste de estos modelos y los compararemos entre sí, mostrando de esta forma cómo pueden modelizarse datos longitudinales de forma efectiva y simple. En estos ejemplos, los distintos modelos alternativos, especialmente los modelos antedependientes, fueron superiores a los modelos de coeficientes aleatorios.

Palabras clave: Antedependencia, modelos Arima, AIC, BIC, estructuras de covarianza, máxima verosimilitud residual, modelos mixtos.



SORT, vol. 27, 2, pp-175-190, 2003

OS

A posteriori disclosure risk measure for tabular data based on conditional entropy

A. Oganian and J. Domingo-Ferrer

Universitat Rovira i Virgili, Dept. of Computer Eng. and Maths.

Statistical database protection, also known as Statistical Disclosure Control (SDC), is a part of information security which tries to prevent published statistical information (tables, individual records) from disclosing the contribution of specific respondents. This paper deals with the assessment of the disclosure risk associated to the release of tabular data. So-called sensitivity rules are currently being used to measure the disclosure risk for tables. These rules operate on an a priori basis: the data are examined and the rules are used to decide whether the data can be released as they stand or should rather be protected. In this paper, we propose to complement a priori risk assessment with a posteriori risk assessment in order to achieve a higher level of security, that is, we propose to take the protected information into account when measuring the disclosure risk. The proposed a posteriori disclosure risk measure is compatible with a broad class of disclosure protection methods and can be extended for computing disclosure risk for a set of linked tables. In the case of linked table protection via cell suppression, the proposed measure allows detection of secondary suppression patterns which offer more protection than others.

Keywords: Statistical disclosure control, statistical databases, tabular data, security.



QÜESTIÓ, vol. 23, 2, pp. 239-259, 1999

MA

Efecto del número de indicadores por factor sobre la identificación y estimación en modelos aditivos de análisis factorial confirmatorio***Effects of number of indicators per factor on identification and estimation of confirmatory factor analysis models***

A. Oliver, J. M. Tomás y M. Hontangas

Àrea de Metodologia de les Ciències del Comportament. Facultat de Psicologia. Universitat de València.

La matriz multirrasgo-multimétodo (MRMM) es un diseño de investigación de larga tradición en Psicología. Las técnicas de análisis de datos adecuadas para una correcta extracción de conclusiones han estado sujetas a controversia. Parece, no obstante, que diversos modelos de análisis factorial confirmatorio resultan muy adecuados. De entre los diversos modelos, dos de ellos han recibido gran atención, el modelo completo, que apareció primero en la literatura, y el de unicidades correlacionadas, que parece una alternativa razonable a los problemas que aparecen en el primero. Los resultados de ambos modelos en la literatura se refieren a situaciones con un solo indicador por combinación rasgo-método. La presente investigación simula datos de matrices MRMM para múltiples indicadores por combinación rasgo-método y somete a prueba la adecuación de las estimaciones de ambos modelos. Los resultados apuntan a un mejor comportamiento del modelo completo, si bien los sesgos, aunque triviales en cuantía, aumentan conforme aumenta la correlación entre métodos.

Palabras clave: Estimación, sesgo, análisis factorial confirmatorio (AFC), matriz MRMM, número de indicadores, correlación entre métodos.



QÜESTIÓ, vol. 25, 2, pp. 337-363, 2001

BI

Análisis de duración mediante un modelo lineal generalizado semiparamétrico***Lifetime data analysis using a semiparametric generalized linear model***

J. Orbe

Departamento de Econometría y Estadística (E.A. III). Universidad del País Vasco/Euskal Herriko Unibertsitatea.

Aitkin y Clayton (1980) proponen el análisis de modelos de duración mediante modelos lineales generalizados. En este trabajo extendemos esta metodología permitiendo que el efecto de alguna de las variables explicativas pueda no ser especificado. Así, el modelo propuesto es un modelo lineal generalizado semiparamétrico, con una componente paramétrica donde se especifica la forma funcional concreta del efecto de las variables explicativas sobre la duración, y una componente no paramétrica donde recogemos el efecto de una variable explicativa sin asumir forma funcional alguna. Desarrollaremos el proceso de estimación así como un procedimiento bootstrap para realizar inferencia. Como aplicación, analizaremos con la metodología propuesta el tiempo de supervivencia para una muestra de pacientes diagnosticados de SIDA.

Palabras clave: Modelos de duración, censura, bootstrap, modelos lineales generalizados, estimación semiparamétrica.

○○○

SORT, vol. 30, 2, pp. 171-192, 2006

RE

Influence diagnostics in exponentiated-Weibull regression models with censored data

E. M. M. Ortega*, V. G. Cancho** and H. Bolfarine***

* ESALQ, University of Sao Paulo.

** ICMC, University of Sao Paulo.

*** IME, University of Sao Paulo.

Diagnostic methods have been an important tool in regression analysis to detect anomalies, such as departures from the error assumptions and the presence of outliers and influential observations with the fitted models. The literature provides plenty of approaches for detecting outlying or influential observations in data sets. In this paper, we follow the local influence approach (Cook 1986) in detecting influential observations with exponentiated-Weibull regression models. The relevance of the approach is illustrated with a real data set, where it is shown that by removing the most influential observations, there is a change in the decision about which model fits the data better.

Keywords: Exponentiated-Weibull distribution, censored data, local influence, influence diagnostic, survival data.

○○○

QÜESTIIÓ, vol. 24, 1, pp. 55-82, 2000

CO

Diseño de algoritmos para el problema del transporte escolar. Aplicación en la provincia de Burgos

Algorithm design for school transportation. Application to Burgos province

J. A. Pacheco, A. Aragón, y C. Delgado

Departamento de Economía Aplicada. Universidad de Burgos.

La problemática del transporte escolar es en Burgos especialmente significativa al ser una provincia extensa con muchos núcleos de población muy dispersos y poco poblados. En este trabajo se describen las aportaciones realizadas por los autores para dar solución a dicho problema, a través de técnicas que den soluciones lo más racionales posibles. En este sentido, hay que indicar que el término de racionalidad no sólo hace referencia a la minimización del coste total del transporte, sino también al cuidado de determinados aspectos como el tiempo que permanecen los alumnos en el vehículo, la elección de carreteras cómodas, etc., en especial a partir de determinados acontecimientos recientes. Asimismo, se muestran los resultados obtenidos con los datos del actual curso.

Palabras clave: Transporte escolar, VRP, VRPTW, búsqueda tabú.

○○○

SORT, vol. 28, 2, pp. 125-160, 2004

SE

Robust estimation and forecasting for beta-mixed hierarchical models of grouped binary data

M. Pashkevich and Y. S. Kharin

Faculty of Applied Mathematics and Computer Science. Belarusian State University. Minsk.

The paper focuses on robust estimation and forecasting techniques for grouped binary data with misclassified responses. It is assumed that the data are described by the beta-mixed hierarchical model (the beta-binomial or the beta-logistic), while the misclassifications are caused by the stochastic additive distortions of binary observations. For these models, the effect of ignoring the misclassifications is evaluated and expressions for the biases of the method-of-moments estimators and maximum likelihood estimators, as well as expressions for the increase in the mean square error of forecasting for the Bayes predictor are given. To compensate the misclassification effects, new consistent estimators and a new Bayes predictor, which take into account the distortion model, are constructed. The robustness of the developed techniques is demonstrated via computer simulations and a real-life case study.

Keywords: Grouped binary data, distortions, hierarchical models, beta-binomial, beta-logistic, robust, estimation, forecasting.



QÜESTIÓ, vol. 24, 3, pp. 425-448, 2000

PD

Variables finitas condicionalmente especificadas***Finite conditionally specified variables***

R. Pérez-Villalta

Universidad de Sevilla. Facultad de Económicas y Empresariales. Dpto. de Economía Aplicada.

En este trabajo se estudia la existencia y unicidad de vectores bidimensionales de variables discreta con recorrido finito, cuando se fijan sus distribuciones condicionadas. Para ello, tras repasar la literatura existente sobre el tema, proporcionamos diversos resultados que relacionan diversos temas de álgebra matricial, especialmente la descomposición singular, con el problema que nos ocupa.

Palabras clave: Distribuciones condicionadas, especificación condicional, rango de una matriz, descomposición singular, cadenas de Markov.



QÜESTIIÓ, vol. 26, 3, pp. 461-482, 2002

TS

Aplicación de redes neuronales artificiales a la previsión de series temporales no estacionarias o no invertibles

The application of artificial neural networks to forecasting non-stationary or non-invertible time series

R. Pino, D. de la Fuente, J. Parreño y P. Priore
Dpto. Admon. de Empresas y Contabilidad. Escuela Politécnica de Ingenieros de Gijón. Universidad de Oviedo.

En los últimos tiempos se ha comprobado un aumento del interés en la aplicación de las Redes Neuronales Artificiales a la previsión de series temporales, intentando explotar las indudables ventajas de estas herramientas. En este artículo se calculan previsiones de series no estacionarias o no invertibles, que presentan dificultades cuando se intentan pronosticar utilizando la metodología ARIMA de Box-Jenkins. Las ventajas de la aplicación de redes neuronales se aprecia con más claridad, cuando se trata de pronosticar sistemas multivariantes no estacionarios.

Palabras clave: Series temporales, previsión, Redes Neuronales Artificiales.



QÜESTIIÓ, vol. 22, 1, pp. 157-171, 1998

OS

The analysis of seasonality in economic statistics: a survey of recent developments

C. Planas
Eurostat.

This article describes the EUROSTAT activities in the field of seasonal adjustment and trend extraction in economic time series. They follow a working program which has been set up during 1995. The attention focuses on X12-REGARIMA (X12 in short), a last update of the X11-family from the Bureau of the Census (see Findley *and al.*, 1996), and on SEATS-TRAMO (see Gomez and Maravall, 1996) which implements the ARIMA-modelbased approach to decompose time series. Three main directions are currently followed: evaluation and comparison of these two methods, construction of a software embodying and interfacing X12 and SEATS-TRAMO, and training in applied time series analysis. The preliminary results which have been obtained on the difficult task of comparing both methods are discussed and the design of the software in construction is presented.

Keywords: Seasonal adjustment, signal extraction, X-11, unobserved components, ARIMA Models, Wiener-Kolmogorov filter.



QÜESTIÓ, vol. 23, 3, pp. 543-558, 1999

OS

Diseño de muestras en encuestas de población y hogares***Sampling design for Household Surveys***

J. Porras Puga
Instituto Nacional de Estadística (INE).

En este artículo se presenta el esquema general del diseño muestral que el Instituto Nacional de Estadística (INE) utiliza en las encuestas dirigidas a la población. Este diseño se conoce como Encuesta General de Población y fue implantado en el año 1971, aunque desde entonces ha sufrido ligeros cambios en cuanto a la periodicidad, criterios de estratificación, tamaño, etc. Se utiliza tanto para las encuestas de tipo continuo (encuestas coyunturales) como de tipo esporádico (encuestas estructurales).

Palabras clave: Marco, estratificación, probabilidad proporcional, estimador de razón.



SORT, vol. 27, 1, pp. 65-78, 2003

BI

Aspects of the Analysis of Multivariate Failure Time Data

R. L. Prentice* and J. D. Kalbfleisch**
* Fred Hutchinson Cancer Research Center.
** University of Michigan.

Multivariate failure time data arise in various forms including recurrent event data when individuals are followed to observe the sequence of occurrences of a certain type of event; correlated failure time when an individual is followed for the occurrence of two or more types of events for which the individual is simultaneously at risk, or when distinct individuals have dependent event times; or more complicated multistate processes when individuals may move among a number of discrete states over the course of a follow-up study and the states and associated sojourn times are recorded. Here we provide a critical review of statistical models and data analysis methods for the analysis of recurrent event data and correlated failure time data. This review suggests a valuable role for partially marginalized intensity models for the analysis of recurrent event data, and points to the usefulness of marginal hazard rate models and nonparametric estimates of pairwise dependencies for the analysis of correlated failure times. Areas in need of further methodology development are indicated.

Keywords: Correlated failure times, independent censoring, marginal models, survivor function estimation, recurrent events.



QÜESTIIÓ, vol. 22, 1, pp. 69-82, 1998

MA

¿Cuántos clusters hay en una población?***How many clusters are there in a population?***

J. J. Prieto Martínez
Universidad Carlos III de Madrid.

Sea una población cerrada formada por un número desconocido K y finito de clusters. El método bootstrap es utilizado para estimar el número de clusters que constituyen una población. Se propone un estimador para K , el cual es ajustado y corregido por su sesgo estimado mediante el método bootstrap de Efron (1979). La varianza del “estimador bootstrap” se calcula por el método jackknife agrupado. Mediante simulación, el estimador es comparado con el de Bickel y Yahav (1985).

Palabras clave: Número de clusters, Bootstrap, Jackknife agrupado.

○○○

QÜESTIIÖ, vol. 22, 2, pp. 291-309, 1998

MA

Estimación del número de clusters en una población aplicando el jackknife generalizado***Estimation of the number of cluster in a population through the generalized jackknife***

J. J. Prieto Martínez
Dpto. de Estadística y Econometría. Universidad Carlos III de Madrid.

Sea una población constituida por un número desconocido de clusters. Este trabajo desarrolla una secuencia finita de estimadores no paramétricos para el número de clusters, basándose en el método jackknife generalizado. Estos estimadores resultan ser una combinación lineal de las frecuencias de observación de cada cluster. Se propone entonces un procedimiento de selección para elegir el más apropiado. La técnica es aplicada a un conjunto de datos reales procedentes de un estudio de captura de especies de una población. Además, se lleva a cabo un estudio de simulación para investigar su comportamiento.

Palabras clave: Número de clusters, jackknife generalizado.

○○○

QÜESTIÓ, vol. 22, 3, pp. 417-441, 1998

SE

Una distribución asintótica para un estimador natural del número de clusters en una población***A little law for a natural estimator of number of clusters in a population***

J. J. Prieto Martínez

Dpto. de Estadística y Econometría. Universidad Carlos III de Madrid.

Un estimador natural, \hat{K} , es propuesto para estimar el número de clusters K existentes en una población heterogénea. Una ley límite normal es rigurosamente probada para dicho estimador. La demostración utiliza un método de Holst (1979). Un ejemplo para un conjunto de datos reales y un estudio realizado por simulación es presentado para el estimador propuesto.

Palabras clave: Clusters, población heterogénea, ley límite normal.



QÜESTIÓ, vol. 25, 3, pp. 523-549, 2001

AO

Secuenciación dinámica de sistemas de fabricación flexible mediante aprendizaje automático: análisis de los principales sistemas de secuenciación existentes***Dynamic scheduling of flexible manufacturing systems through machine learning: an analysis of the main scheduling systems***

P. Priore, D. De la Fuente, J. Puente y A. Gómez

Dpto. de Administración de Empresas y Contabilidad. Escuela Técnica Superior de Ingenieros Industriales e Informáticos de Gijón. Universidad de Oviedo.

Una forma habitual de secuenciar de modo dinámico los trabajos en los sistemas de fabricación es mediante el empleo de reglas de secuenciación. Sin embargo, el problema que presenta este método es que el comportamiento del sistema de fabricación dependerá de su estado, y no existe una regla que supere a las demás en todos los posibles estados que puede presentar el sistema de fabricación. Por lo tanto, sería interesante usar en cada momento, la regla más adecuada. Para lograr este objetivo, se pueden utilizar sistemas de secuenciación que emplean aprendizaje automático que permiten, analizando el comportamiento previo del sistema de fabricación (ejemplos de entrenamiento), obtener el conocimiento necesario para determinar la regla de secuenciación más apropiada en cada instante. En el presente trabajo se realiza una revisión de los principales sistemas de secuenciación, existentes en la literatura, que utilizan aprendizaje automático para variar de forma dinámica la regla de secuenciación empleada en cada momento.

Palabras clave: Secuenciación dinámica, aprendizaje automático, sistemas de fabricación flexible, reglas de secuenciación, simulación.



SORT, vol. 31, 1, pp. 45-54, 2007

SI

Goodness of fit tests for the skew-Laplace distribution

P. Puig* and M. A. Stephens**

* Departament de Matemàtiques, Universitat Autònoma de Barcelona.

** Department of Statistics & Actuarial Science, Simon Fraser University, B.C. Canada.

The skew-Laplace distribution is frequently used to fit the logarithm of particle sizes and it is also used in Economics, Engineering, Finance and Biology. We show the Anderson-Darling and Cramér-von Mises goodness of fit tests for this distribution.

Keywords: Maximum likelihood estimator, non-regular models, Anderson-Darling statistic, Cramér-von Mises statistic.

○○○

QÜESTIÓ, vol. 25, 2, pp. 211-223, 2001

PD

The proportional likelihood ratio order and applications

H. M. Ramos Romero and M. A. Sordo Díaz

Departamento de Estadística e I. O. Universidad de Cádiz.

H. M. Ramos Romero and M. A. Sordo Díaz In this paper, we introduce a new stochastic order between continuous nonnegative random variables called the PLR (proportional likelihood ratio) order, which is closely related to the usual likelihood ratio order. The PLR order can be used to characterize random variables whose logarithms have log-concave (log-convex) densities. Many income random variables satisfy this property and they are said to have the IPLR (increasing proportional likelihood ratio) property (DPLR property). As an application, we show that the IPLR and DPLR properties are sufficient conditions for the Lorenz ordering of truncated distributions.

Keywords: Likelihood ratio order, proportional likelihood ratio order, ILR, DLR, IPLR, DPLR, log-concave density function, Lorenz order, truncated distributions.

○○○

QÜESTIÓ, vol. 26, 1-2, pp. 15-28, 2002

PD

Characterizations of inequality orderings by means of dispersive orderings

H. M. Ramos and M. A. Sordo

Departamento de Estadística e I. O. Universidad de Cádiz.

The generalized Lorenz order and the absolute Lorenz order are used in economics to compare income distributions in terms of social welfare. In Section 2, we show that these orders are equivalent to two stochastic orders, the concave order and the dilation order,

which are used to compare the dispersion of probability distributions. In Section 3, a sufficient condition for the absolute Lorenz order, which is often easy to verify in practice, is presented. This condition is applied in Section 4 to the ordering of generalized gamma distributions with different parameters.

Keywords: Generalized Lorenz order, absolute Lorenz order, concave order, dilation order.



QÜESTIÓ, vol. 26, 3, pp. 429-441, 2002

MA

Implementación del cálculo de polinomios zonales i aplicaciones en análisis multivariante

Implementing calculus of zonal polynomials and applications in multivariate analysis

J. Rodríguez-Avi, A. J. Sáez-Castillo y A. Conde-Sánchez
Dep. de Estadística e I.O. Universidad de Jaén.

En este trabajo se describe la implementación de un algoritmo para el cálculo de polinomios zonales, así como dos aplicaciones explícitas de éstos en el ámbito del análisis multivariante. Concretamente, esta implementación permite obtener resultados de sumación aproximados para funciones hipergeométricas de argumento matricial que, a su vez, pueden utilizarse en la génesis de distribuciones multivariantes discretas con frecuencias simétricas. De igual forma, se pone en práctica un conocido resultado teórico que caracteriza la distribución de la menor raíz característica de una matriz aleatoria con distribución de Wishart.

Palabras clave: Polinomios zonales, distribuciones discretas, distribución de Wishart, funciones hipergeométricas de argumento matricial.



QÜESTIÓ, vol. 24, 3, pp. 449-491, 2000

TS

Análisis de detección de raíces unitarias en series de tiempo. Un enfoque metodológico con tests no similares

Testing for unit roots in time series: An analysis with non similar tests

J. A. Roldán Casas y R. Dios Palomares
Universidad de Córdoba. Dpto. de Estadística e Investigación Operativa. E.S.I.A y de Montes.

El presente artículo recoge los resultados de una investigación llevada a cabo con el fin de analizar, desde la perspectiva de la no similaridad, las distribuciones de los distintos estadísticos planteados por Dickey y Fuller para contrastar la presencia de raíz unitaria. Asimismo, se definen zonas de rechazo y aceptación de las hipótesis nulas para cada estadístico, considerando las distintas distribuciones del mismo, y se estudian las situaciones

con las que nos podemos encontrar de cara a deducir una pauta de comportamiento que minimice el riesgo de error. Se presenta un ejemplo en el que se pone de manifiesto que la no similaridad de los contrastes tradicionales de raíz unitaria nos puede llevar a tomar una decisión equivocada. Teniendo en cuenta dicha no similaridad, se propone una estrategia de contraste secuencial para resolver situaciones de indecisión acerca del rechazo o no de la existencia de raíz unitaria. Finalmente, se lleva a cabo un experimento Monte Carlo para determinar empíricamente la probabilidad de que el valor calculado de los estadísticos implicados en el contraste caigan en las diferentes zonas planteadas.

Palabras clave: Raíz unitaria, tendencia estocástica, test no similar, test de Dickey-Fuller, experimento Monte Carlo.

○○○

SORT, vol. 27, 2, 139-152, 2003

GT

Partial cooperation and convex sets

H. M. Ramos and M. A. Sordo
Departamento de Estadística e I. O. Universidad de Cádiz.

The generalized Lorenz order and the absolute Lorenz order are used in economics to compare income distributions in terms of social welfare. In Section 2, we show that these orders are equivalent to two stochastic orders, the concave order and the dilation order, which are used to compare the dispersion of probability distributions. In Section 3, a sufficient condition for the absolute Lorenz order, which is often easy to verify in practice, is presented. This condition is applied in Section 4 to the ordering of generalized gamma distributions with different parameters.

Keywords: Generalized Lorenz order, absolute Lorenz order, concave order, dilation order.

○○○

SORT, vol. 27, 2, 139-152, 2003

GT

Partial cooperation and convex sets

J. E. Romero García, J. J. López Vázquez
Facultad de Ciencias Económicas y Empresariales. Universidad de Sevilla.

We consider games of transferable utility, those that deal with partial cooperation situations, made up of coalition systems, in which every unit coalition is feasible and every coalition of players can be expressed as a disjoint union of maximal feasible coalitions. These systems are named partition systems and cause restricted games. To sum up, we study feasible coalition systems defined by a partial order designed for a set of players and we analyze the characteristics of a feasible coalition system developed from a family of convex sets.

Keywords: Cooperative games, partial cooperation, convex sets.

○○○

QÜESTIÓ, vol. 24, 2, pp. 367-389, 2000

OS

Winsorització de la despesa turística a Catalunya***Winsorization of tourism expenditure in Catalonia***

M. Sáez, M. A. Barceló, C. Saurina i G. Coenders
 Departament d'Economia. Universitat de Girona.

En el marc d'actuació de l'Institut d'Estadística de Catalunya (Idescat), el sector turístic apareix com una línia prioritària, tant des de l'àmbit de la producció estadística com des de la reflexió teòrica, i dins aquest context es pot destacar la despesa turística com a variable certament estratègica, quan es tracta d'aprofundir en temes tan determinants com poden ser el de l'impacte del turisme en el creixement econòmic o el de la comptabilitat satèl·lit del sector turístic. Això justifica abastament l'interès de l'Idescat en el desenvolupament de millores metodològiques en aquesta unitat d'anàlisi. En el present article es proposa un mètode per a imputar les dades mancants i corregir els errors sistemàtics de les enquestes destinades a mesurar la despesa dels viatges turístics. La complexitat de la despesa turística fa que aquests errors siguin sobretot per omissió. El mètode proposat consisteix en winsoritzar asimètricament els residus d'un model de regressió que inclou les principals característiques del viatge realitzat com a variables predictores de la despesa diària per persona. Per a l'estimació s'empra iterativament el mètode de mínims quadrats, cosa que es pot considerar una variant de l'algorisme EM. Els errors estàndard es calculen per remostreig amb el mètode jackknife. El procediment s'il·lustra amb dades dels viatges a Catalunya realitzats per catalans, altres ciutadans de l'Estat espanyol i estrangers durant l'any 1998.

Paraules clau: Winsorització, regressió robusta, jackknife, valors atípics, dades mancants, despesa turística.

○○○

QÜESTIÓ, vol. 22, 2, pp. 243-252, 1998

ST

On some strategies using auxiliary information for estimating finite population mean

L. N. Sahoo*, J. Sahoo** and M. Ruiz Espejo***

* Department of Statistics, Utkal University, India.

** Department of Statistics, Orissa University of Agriculture and Technology, India.

*** Departamento de Economía Aplicada Cuantitativa. Universidad Nacional de Educación a Distancia (UNED).

This paper presents an empirical investigation of the performances of five strategies for estimating the finite population mean using parameters such as mean or variance or both of an auxiliary variable. The criteria used for the choices of these strategies are bias, efficiency and approach to normality (asymmetry).

Keywords: Auxiliary information, empirical study, finite population mean, sampling.

○○○

QÜESTIÓ, vol. 22, 2, pp. 339-361, 1998

OS

Mètode per a l'estimació de migracions en àrees petites a partir de dades incompletes***Method for estimating small area migration from incomplete data***

J. A. Sánchez Cepeda
Institut d'Estadística de Catalunya.

La disposició de dades de migració per a àrees petites s'emmarca dins el projecte d'estimacions de població, consistent en determinar la població recent de Catalunya per comarques i municipis majors de 45.000 habitants, desagregada per sexe i edat. La no disposició de dades amb aquest nivell de desagregació per alguns anys ha portat a la implementació d'un mètode per a la seva estimació. Aquest treball descriu el mètode d'estimació dels efectius migratoris per àrea geogràfica, sexe i edat que s'aplica a l'Institut d'Estadística de Catalunya (Idescat). El mètode consisteix en calcular per separat la intensitat i el perfil tant de l'emigració com de la immigració. El principal problema en l'estimació del nivell o intensitat migratòria és haver de treballar amb àrees menors de 45.000 habitants. En el càlcul del perfil migratori es redueix el nombre de perfils dels 62 inicials a 6, que constitueixen els patrons migratoris observats a Catalunya. Es comenta la tipologia dels patrons migratoris, en base als paràmetres dels mateixos determinats pel model. Finalment, es calcula el nombre d'emigrants i immigrants a partir de la intensitat i el perfil migratori estimats.

Paraules clau: Àrea petita, model migratori, anàlisi cluster, ajust no lineal.

○○○

QÜESTIÓ, vol. 24, 3, pp. 555-576, 2000

BI

Cálculo de probabilidades en el análisis de perfiles genéticos compatibles***Calculation of evidence in the analysis of genetic compatible profiles***

M. Sánchez García y P. Cuesta Alvaro
Departamento de Estadística. Facultad de Medicina. Universidad Complutense de Madrid.

Un hecho delictivo ha sido cometido por una o más personas. Para descubrir el número de presuntos delincuentes, se determina el perfil genético de la evidencia forense encontrada en el lugar del delito, mediante el estudio de marcadores de tipo STR. En el presente artículo se desarrollan diversos algoritmos. En el primero se considera la probabilidad de que n personas elegidas aleatoriamente de la población de referencia contengan un perfil genético compatible, en el segundo se calcula la evidencia del material genético y en el tercero se calcula el mínimo número de presuntos culpables seleccionado de un conjunto compatible. Finalmente, se analizan las evidencias en diversas situaciones particulares.

Palabras clave: STR, alelo, principio de inclusión-exclusión, multigrafo, problema de recubrimiento.

○○○

QÜESTIÓ, vol. 22, 1, pp. 83-99, 1998

RE

Una nota sobre el contraste de relaciones de cointegración entre índices de precios***Testing Cointegration Relationships between price indexes. A note***

A. Sansó Rosselló, M. Artís Ortuño i J. Suriñach Caralt

Departament d'Econometria, Estadística i Economia Espanyola. Universitat de Barcelona.

El contraste de relaciones de equilibrio usando técnicas de cointegración, entre variables económicas en general y entre índices de precios en particular, ha recibido elevada atención en los últimos años. De entre estas últimas relaciones destacar el contraste de la hipótesis de la Paridad de Poder de Compra (PPP). Johnson (1991) desarrolla un contraste para tratar de explicar los rechazos que de dicha hipótesis se producían. El supuesto de qué parte es qué, aun cuando la hipótesis PPP se cumpla para los bienes individuales, cambios relativos de precios en los índices unido a diferentes ponderaciones, pueden provocar dichos rechazos. En artículo que se presenta se realiza una extensión y crítica del trabajo citado. Respecto a las extensiones, se amplía el análisis a índices de Laspeyres, a la desagregación del índice en subíndices de grupos de gasto y al análisis de convergencia en tasas de inflación. Respecto a las críticas, se muestran las diferentes interpretaciones del contraste propuesto, las limitaciones en cuanto a su aplicación y, en un intento de solucionar éstas, las pobres conclusiones que pueden extraerse de analizar los cambios relativos entre índices de grupo de gasto. De los desarrollos propuestos en la nota se desprende que el contraste de relaciones de equilibrio entre índices de precios ponderados presenta la dificultad de que un rechazo de dicha relación no puede interpretarse directamente como un rechazo de la hipótesis económica subyacente (PPP, convergencia en tasas de inflación, etc.).

Palabras clave: Índices de precios, Cointegración, PPP, Convergencia.

SORT, vol. 28, 2, pp. 215-230, 2004

OS

A comparative study of small area estimators

L. Santamaría*, D. Morales* and I. Molina**

* Centro de Investigación Operativa. Universidad Miguel Hernández de Elche.

** Departamento de Estadística. Universidad Carlos III de Madrid.

It is known that direct-survey estimators of small area parameters, calculated with the data from the given small area, often present large mean squared errors because of small sample sizes in the small areas. Model-based estimators borrow strength from other related areas to avoid this problem. How small should domain sample sizes be to recommend the use of model-based estimators? How robust small area estimators are with respect to the rate sample size/number of domains? To give answers or recommendations about the questions above, a Monte Carlo simulation experiment is carried out. In this simulation study, model-based estimators for small areas are compared with some standard design-based estimators. The simulation study starts with the construction of an artificial population data file, imitating a census file of an Statistical Office. A stratified random design is used to draw samples from the artificial population. Small area estimators of the mean of a continuous variable

are calculated for all small areas and compared by using different performance measures. The evolution of this performance measures is studied when increasing the number of small areas, which means to decrease their sizes.

Keywords: Small area estimation, eblup estimators, sampling designs, mixed linear models.



QÜESTIÓ, vol. 26, 3, pp. 517-536, 2002

NF

Un algoritmo para el problema de biflujo máximo simétrico no dirigido

An algorithm for the undirected symmetric maximum biflow problema

A. Sedeño-Noda y C. González-Martín

Departamento de Estadística, Investigación Operativa y Computación (DEIOC). Universidad de La Laguna.

En este trabajo proponemos un algoritmo de $O(nm \log U)$ para resolver el problema de biflujo máximo simétrico en una red no dirigida. Para resolver este problema se introduce un cambio de variable que permite dividir el problema original en dos problemas de flujo máximo. De esta manera se obtiene un algoritmo sencillo y eficiente donde se utilizan las herramientas computacionales propias de la resolución del clásico problema de maximizar un único flujo.

Palabras clave: Problema de biflujo máximo simétrico, algoritmo de escalado en las capacidades, redes no dirigidas.



SORT, vol. 31, 1, pp. 75-96, 2007

SE

Nonparametric bivariate estimation for successive survival times

C. Serrat and G. Gómez

Departament de Matemàtica Aplicada I. Universitat Politècnica de Catalunya.

Several aspects of the analysis of two successive survival times are considered. All the analyses take into account the dependent censoring on the second time induced by the first. Three nonparametric methods are described, implemented and applied to the data coming from a multicentre clinical trial for HIV-infected patients. Visser's and Wang and Wells methods propose an estimator for the bivariate survival function while Gómez and Serrat's method presents a conditional approach for the second time given the first. The three approaches are compared and discussed at the end of the paper.

Keywords: AIDS, Conditional survival, Dependent censoring, Inverse probability of censoring weighted estimators, Successive survival times.



QÜESTIÓ, vol. 22, 3, pp. 403-416, 1998

ST

Almost unbiased ratio and product-type estimators in systematic sampling

R. Singh* and H. P. Singh**

* Prestige Institute of Management, Mandasaur.

** School of Studies in Statistics, Vikram University, Ujjain-456010, M.P., India.

In this paper we have suggested almost unbiased ratio-type and product-type estimators for estimating the population mean \bar{Y} of the study variate y using information on an auxiliary variate x in systematic sampling. The variance expressions of the suggested estimators have been obtained and compared with usual unbiased estimator \bar{y}^* , Swain's (1964) ratio estimator \bar{y}_R^* and Shukla's product estimator \bar{y}_P^* . It has been shown that the proposed estimators are more efficient than usual unbiased estimator \bar{y}^* , ratio estimator \bar{y}_R^* and product estimator \bar{y}_P^* . An empirical study is carried out to demonstrate the superiority of the constructed estimators over the estimators \bar{y}^* , \bar{y}_R^* and \bar{y}_P^* .

Keywords: Almost unbiased ratio and product-type estimators, Auxiliary information, Bias, Variance, Systematic sampling.



QÜESTIÓ, vol. 22, 3, pp. 471-489, 1998

NF

Tour eulerià sense girs en U en un graf orientat simple***Eulerian tour without U-turns in a simple digraph***

D. Soler Fernández

Departament de Matemàtica Aplicada. Universitat Politècnica de València.

Si $G = (V, A)$ un graf orientat eulerià simple, hi estudiem la recerca d'un tour eulerià sense girs en U, és a dir, sense recórrer consecutivament parells d'arcs (u, v) , (v, u) , $u, v \in V$. Desconeguda la complexitat d'aquest problema, generalitzem un resultat d'un cas particular resolt en temps polinomial, proporcionant una condició sota la qual es pot construir en temps polinomial un tour eulerià sense girs en U sobre G . Aquesta condició es basa, a més a més, en l'eliminació de vèrtexs candidats a contenir girs en U en un tour eulerià amb mínim nombre d'ells.

Paraules clau: Gir en U, graf orientat U-eulerià, H-subgraf.



SORT, vol. 31, 2, pp. 169-180, 2007

GE

Poverty measures and poverty orderings

M. A. Sordo*, H. M. Ramos* and C. D. Ramos**

* Departamento de Estadística e Investigación Operativa, Facultad de Ciencias Económicas y Empresariales, Universidad de Cádiz.

** Departamento de Estadística e Investigación Operativa, E.U.E. Empresariales, Universidad de Cádiz.

We examine the conditions under which unanimous poverty rankings of income distributions can be obtained for a general class of poverty indices. The “per-capita income gap” and the Shorrocks and Thon poverty measures are particular members of this class. The conditions of dominance are stated in terms of comparisons of the corresponding TIP curves and areas.

Keywords: Poverty measure, poverty ordering, TIP curve.



SORT, vol. 30, 2, pp. 205-218, 2006

BI

Univariate Parametric Survival Analysis using GS-distributions

A. Sorribas, J. M. Muiño, M. Rué and J. Fibla

Departament de Ciències Mèdiques Bàsiques, Institut de Recerca Biomèdica de Lleida (IR-BLLEIDA), Universitat de Lleida.

The GS-distribution is a family of distributions that provide an accurate representation of any unimodal univariate continuous distribution. In this contribution we explore the utility of this family as a general model in survival analysis. We show that the survival function based on the GS-distribution is able to provide a model for univariate survival data and that appropriate estimates can be obtained. We develop some hypotheses tests that can be used for checking the underlying survival model and for comparing the survival of different groups.

Keywords: Univariate distributions, survival analysis, Kaplan-Meier.



SORT, vol. 29, 2, pp. 201-216, 2005

SI

On the role played by the fixed bandwidth in the Bickel-Rosenblatt goodness-of-fit test

C. Tenreiro

Universidade de Coimbra.

For the Bickel-Rosenblatt goodness-of-fit test with fixed bandwidth studied by Fan (1998) we derive its Bahadur exact slopes in a neighbourhood of a simple hypothesis $f = f_0$ and we use them to get a better understanding on the role played by the smoothing parameter in the detection of departures from the null hypothesis. When f_0 is a univariate normal distribution and we take for kernel the standard normal density function, we compute these slopes for a set of Edgeworth alternatives which give us a description of the test properties in terms of the bandwidth h . A simulation study is presented which indicates that finite sample properties are in good accordance with the theoretical properties based on Bahadur local efficiency. Comparisons with the quadratic classical EDF tests lead us to recommend a test based on a combination of bandwidths in alternative to Anderson-Darling or Cramér-von Mises tests.

Keywords: Goodness-of-fit test, kernel density estimator, Bahadur efficiency.



SORT, vol. 29, 2, pp. 217-234, 2005

BS

On sequential and fixed designs for estimation with comparisons and applications

M. Terbeche*, B. O. Oluyede** and A. Barbour***

* University of Oran.

** Georgia Southern University.

*** U.A.E. University.

A fully sequential approach to the estimation of the difference of two population means for distributions belonging to the exponential family of distributions is adopted and compared with the best fixed design. Results on the lower bound for the Bayes risk due to estimation and expected cost are presented and shown to be of first order efficiency. Applications involving the Poisson and exponential distributions with gamma priors as well as the Bernoulli distribution with beta priors are given. Finally, some numerical results are presented.

Keywords: Estimation, Loss function, Prior distribution, Sequential design.



QÜESTIÓ, vol. 23, 3, pp. 491-505, 1999

OS

Utilisation d'informations auxiliaires dans les enquêtes par sondage

Use of auxiliary information in survey sampling

Y. Tillé

Laboratoire de Statistique d'Enquête. CREST-ENSAI. École Nationale de la Statistique and de l'Analyse de l'Information. Bruz.

La notion de représentativité est apparue dès la naissance de la théorie des sondages à la fin du dix-neuvième siècle. Pourtant, ce concept qui est appliqué autant aux plans par quotas qu'aux plans probabilistes est largement galvaudé. Après avoir rappelé quelques éléments de l'histoire de la théorie des sondages, nous rappelons quelques techniques de

base de plans aléatoires et à choix raisonnés. Nous montrons ensuite que le concept de plan équilibré permet de lever les ambiguïtés fondamentales de la notion de représentativité.

Mots clés: Sondage, plans équilibrés, représentativité.



SORT, vol. 27, 1, pp. 79-94, 2003

BI

Indirect inference for survival data

B. W. Turnbull and W. Jiang
Cornell University, Ithaca, NY.

In this paper we describe the so-called “indirect” method of inference, originally developed from the econometric literature, and apply it to survival analyses of two data sets with repeated events. This method is often more convenient computationally than maximum likelihood estimation when handling such model complexities as random effects and measurement error, for example; and it can also serve as a basis for robust inference with less stringent assumptions on the data generating mechanism. The first data set concerns recurrence times of mammary tumors in rats and is modeled using a Poisson process model with covariates and frailties. The second data set involves times of recurrences of skin tumors in individual patients in a clinical trial. The methodology is applied in both parametric and semi-parametric regression analyses to accommodate random effects and covariate measurement error.

Keywords: Estimating equations, frailty, hazard rate regression, indirect inference, measurement, error, naive estimators, overdispersion, quasi-likelihood, random effects, robustness.



QÜESTIIÓ, vol. 26, 1-2, pp. 123-138, 2002

TS

Un modelo poissoniano para predecir la matriculación de vehículos en países europeos

A Poissonian model to forecast vehicles matriculations in european countries

M. J. Valderrama, A. M. Aguilera y P. R. Bouzas
Departamento de Estadística e I.O., Facultad de Farmacia, Campus de Cartuja, Universidad de Granada.

En este artículo presentamos una modelización para la matriculación de vehículos en países de Europa mediante un proceso de Poisson Doblemente Estocástico con media aleatoria Normal truncada. Apoyándonos en trabajos previos acerca de este proceso, se amplía el estudio de características de éste. Asimismo, se hace una predicción de este proceso para los años 2000 y 2001.

Palabras clave: Proceso de Poisson Doblemente Estocástico, Distribución Normal truncada, Matriculaciones de Turismos.



QÜESTIÓ, vol. 22, 3, pp. 443-468, 1998

RE

Algunas soluciones aproximadas para diseños split-plot con matrices de covarianza arbitrarias

Some approximate solutions for split-plot designs with arbitrary covariance matrices

G. Vallejo Seco y J. R. Escudero García
Departamento de Psicología. Universidad de Oviedo.

El presente trabajo revisa con cierto detalle diversos tipos de análisis para diseños split-plot que carecen del mismo número de unidades experimentales dentro de cada grupo y en los que se incumple con el supuesto de esfericidad multimuestral. Específicamente, adaptando el enfoque multivariado de aproximar los grados de libertad desarrollado por Johansen (1980) y el procedimiento de la aproximación general mejorada corregida basado en Huynh (1980) se muestra cómo obtener análisis robustos y poderosos a la hora de probar los efectos principales y la interacción, así como hipótesis de comparaciones múltiples relacionadas con estos efectos, tanto si se cuenta con una simple variable dependiente asociada con cada una de las medidas repetidas como si se cuenta con más de una.

Palabras clave: Diseños multivariados de medidas repetidas, esfericidad multimuestral, medias ponderadas y no ponderadas, procedimientos robustos y poderosos.



QÜESTIÓ, vol. 24, 2, pp. 293-313, 2000

MA

Contrastación de hipótesis en diseños multivariados split-plot con matrices de dispersión arbitrarias

Testing of hypotheses in multivariate split-plot designs with arbitrary dispersion matrices

G. Vallejo, J. R. Escudero, A. M. Fidalgo y M.^a P. Fernández
Departamento de Psicología. Universidad de Oviedo.

El presente trabajo examina diversos procedimientos para contrastar hipótesis nulas globales, correspondientes a datos obtenidos mediante diseños multivariados split-plot cuando se incumple el supuesto de homogeneidad de las matrices de dispersión. Un examen de estos procedimientos para un amplio número de variables confirma, por un lado, la robustez del procedimiento multivariado de Welch-James dado por Johansen (1980) para probar el efecto principal de los ensayos y, por otro, la robustez de la generalización multivariada del procedimiento de Brown-Forsythe (1974) para probar la interacción de los grupos x los en-

sayos. Nuestros resultados también ponen de relieve que las diferencias de potencia eran pequeñas en aquellas condiciones en que tanto el procedimiento de Welch-James como de Brown-Forsythe controlaban las tasas de error de Tipo I.

Palabras clave: Diseños multivariados split-plot, matrices de dispersión arbitrarias, procedimientos robustos y potentes.



SORT, vol. 28, 1, pp. 87-108, 2004

BI

Incorporating patients' characteristics in cost-effectiveness studies with clinical trial data. A flexible Bayesian approach

F. J. Vázquez-Polo and M. A. Negrín-Hernández
Dept. of Quantitative Methods. Fac. CC. Económicas. University of Las Palmas de G.C.

Most published research on the comparison between medical treatment options merely compares the results (effectiveness and cost) obtained for each treatment group. The present work proposes the incorporation of other patient characteristics into the analysis. Most of the studies carried out in this context assume normality of both costs and effectiveness. In practice, however, the data are not always distributed according to this assumption. Alternative models have to be developed. In this paper, we present a general model of cost-effectiveness, incorporating both binary effectiveness and skewed cost. In a practical application, we compare two highly-active antiretroviral treatments applied to asymptomatic HIV patients. We propose a logit model when the effectiveness is measured depending on if an initial purpose is achieved. For this model, the measure to compare treatments is the difference in the probability of success. Besides, the cost data usually present a right skewing. We propose the use of the log-transformation to carry out the regression model. The three models are carried out demonstrating the advantages of this modelling. The cost-effectiveness acceptability curve is used as measure to take decisions.

Keywords: Bayesian Analysis, Cost-Effectiveness, Markov Chain Monte Carlo (MCMC), Skewed Cost Distributions.



QÜESTIÓ, vol. 25, 1, pp. 93-117, 2001

OS

Product expenditure patterns in the ECPF survey: an analysis using multiple group latent-variables models

E. Ventura and A. Satorra
Universitat Pompeu Fabra.

Using data from the Spanish household budget survey, we investigate some aspects of household heterogeneity on several product expenditures. We adopt a latent-variable model approach to evaluate the impact of income on expenditures, controlling for the number of members in the family. Two latent factors underlying repeated measures of monetary and

non-monetary income are used as explanatory variables in the expenditure regression equations, thus avoiding possible bias associated to the measurement error in income. The proposed methodology also takes care of the case in which product expenditures exhibit a pattern of infrequent purchases. Multiple-group analysis is used to assess the variation of key parameters of the model across various household typologies. The analysis discloses significant variations across groups on the mean levels of expenditures and on the way income and family size affect expenditures. Asymptotic robust methods are used to account for possible non-normality of the data.

Keywords: Multiple group, latent variables, product expenditures, ECPF.

○○○

QÜESTIÓ, vol. 23, 3, pp. 507-524, 1999

OS

Muestreo y recogida de datos en el análisis de redes sociales

Data collection and sampling in Social Network Analysis

J. M. Verd Pericàs y J. Martí Olivé

Grup d'Estudis Sociològics sobre la Vida Quotidiana i el Treball (QUIT). Departament de Sociologia. Universitat Autònoma de Barcelona.

El artículo revisa las propuestas que dentro de la perspectiva del Análisis de Redes Sociales han realizado diversos autores en relación al muestreo y la recogida de datos. Estos aspectos, resueltos de modo satisfactorio en la perspectiva individualista-atomista, plantean no pocos problemas en la perspectiva de redes sociales. Resulta especialmente problemática la posibilidad de realizar muestras representativas de las relaciones existentes en una población. Aún en el caso de conocer con antelación todos los actores y sus relaciones entre ellos, la elección de una muestra representativa de actores no garantiza una muestra representativa de relaciones. La alternativa tomada en el artículo es la de realizar el análisis de grupos sociales tomados como "poblaciones". Una aplicación práctica de esta última opción es presentada como ejemplo.

Palabras clave: Redes sociales, datos relacionales, muestreo.

○○○

QÜESTIÓ, vol. 24, 1, pp. 189-204, 2000

BI

Comparación de dos tablas demográficas: aproximación a su significación estadística

Comparison of two demographic tables: approximation to their statistical significance

E. J. Veres

Departamento de Economía Aplicada. Facultad de Ciencias Económicas y Empresariales. Universidad de Valencia.

En este trabajo se analiza la significación estadística de la posible igualdad de dos tablas demográficas. Concretamente, se presenta la aplicabilidad de un clásico contraste de los métodos estadísticos –el de homogeneidad de dos distribuciones– para contrastar la hipótesis nula “ H_0 = las dos tablas demográficas son iguales, esto es, responden a una misma estructura del fenómeno demográfico estudiado”, frente a la alternativa que niega la anterior. Ambas tablas se refieren a un único fenómeno demográfico para el mismo ámbito territorial y dos momentos diferentes de tiempo (contraste temporal), o para la misma referencia temporal en dos poblaciones de ámbito territorial distinto (contraste territorial). Se aplica la metodología descrita en dos situaciones: para la comparación de los niveles de mortalidad de dos provincias, y sobre dos tablas-tipo correspondientes a sendos niveles de mortalidad.

Palabras clave: Calendario, contraste de hipótesis, intensidad, tabla demográfica, tablas-tipo de mortalidad, test de la χ^2 .



QÜESTIÓ, vol. 25, 2, pp. 303-334, 2001

OS

Tratamiento de datos territorializados de vivienda en el inventario de capital residencial

Local housing data processing for the perpetual inventory of residencial capital

R. Vergés Escuín

Consejo Superior de Colegios de Arquitectos de España.

Después de plantear la filosofía de la contabilidad de capital fijo, el artículo examina la metodología necesaria para levantar el inventario permanente residencial de cada territorio. Al nivel de flujos, se ordenan y comparan las distintas fuentes de datos sobre edificación que existen en España y se describe el tratamiento requerido por cada una de ellas. Al nivel de stocks, se propone un procedimiento de ajuste y de desglose territorial del parque de viviendas por fecha de construcción. También se desarrolla la metodología para proyectar la supervivencia de los estratos de mismo período de construcción a partir de funciones de agotamiento. Se integra asimismo la variable de suelo urbanizable según el planeamiento vigente como indicador del desarrollo de nuevos estratos en las proyecciones del inventario. Por último se analizan algunos problemas acerca de la utilización de datos sobre clases de vivienda. El artículo concluye con la perspectiva de una valoración del capital fijo residencial al nivel territorial gracias a la emergencia de nueva información acerca de los mercados de vivienda.

Palabras clave: Capital, depreciación, inventario permanente, planeamiento, suelo, vivienda.



QÜESTIÓ, vol. 23, 3, pp. 465-488, 1999

MS

Independencia entre las cuestiones en el análisis factorial de tablas disyuntivas incompletas con preguntas condicionadas***Independence between questions in the factor analysis of incomplete disjunctive tables with conditioned questions***

A. Zárraga y B. Goitisoló

Universidad del País Vasco-Euskal Herriko Unibertsitatea. Dpto. de Economía Aplicada III. Fac. CCEE y EE.

El análisis de correspondencias múltiples (ACM) estudia la relación entre varias variables cualitativas definidas sobre una misma población. Sin embargo, una de las principales fuentes de información son las encuestas donde es frecuente encontrar cierto número de datos ausentes y de preguntas condicionadas. Escofier (Escofier 1981) propone analizar la tabla disyuntiva incompleta sustituyendo la marginal real de la tabla sobre los individuos por una marginal impuesta constante. El análisis de la tabla disyuntiva incompleta está asociado a unas tasas de inercia pequeñas que no deben ser interpretadas como partes de información explicada por los ejes. Se estudiará el caso en el cual las cuestiones son independientes dos a dos y se propondrá una corrección a estas tasas de inercia en el análisis con marginal modificada de una tabla disyuntiva incompleta.

Palabras clave: Análisis de correspondencias múltiples, tabla disyuntiva incompleta, independencia entre variables cualitativas, valores propios, tasas de inercia.

