

Colaboración

Software

Actividades da SGAPEIO

Novas do IGE

Traballos de Estatística e IO

Coñecéndonos

Sabías que?

Editorial

ANTONIO VAAMONDE LISTE – EX-PRESIDENTE DA SGAPEIO

Xubilación de Patricio Sánchez Bello (Uvigo)

O Departamento de Estatística e Investigación Operativa da Universidade de Vigo celebrou recentemente unha homenaxe de despedida ao Profesor Patricio Sánchez Bello, quen acaba de recibir os beneficios da xubilación despois de máis de corenta anos dedicado á docencia Universitaria da Estatística e Investigación Operativa. Os seus compañeiros docentes valoran e aprecian a súa permanente dispoñibilidade, a súa colaboración entusiasta nos proxectos comúns, a súa dedicación sen queixas nin reservas e a eficacia da súa insuperable capacidade de relación, de grande utilidade na resolución dos problemas do día a día na xestión universitaria.



Patricio deixa atrás máis de dúas décadas na Dirección da Escola Universitaria de Estudos Empresariais de Vigo, como Secretario e como Director (nestes últimos anos), co recoñecemento unánime da comunidade universitaria, así como do colectivo empresarial e social de Vigo, recoñecemento especialmente meritorio pola época conflitiva que lle tocou vivir, e pola complexidade dos problemas que tivo que abordar e resolver. Como responsable de deportes da Universidade de Vigo nos primeiros anos da súa creación, é o artífice da posta en marcha de moitas das instalacións, equipos e competicións coas que conta a Universidade, polo que é recordado neste ámbito tamén con cariño e admiración.

Cada afán ten o seu tempo. Unhas tarefas terminan, pero outras continúan ou empezan. Sabemos que Patricio non abandona totalmente a docencia xa que permanece en programas voluntarios de extensión universitaria, nin abandona a investigación no ámbito da Estatística aplicada, dende o grupo de traballo co que mantén os vínculos anteriores con algún proxecto en curso, e dende o Instituto de Estudos Vigueses do que é membro numerario, pero agora pode dedicar máis tempo a outros traballos, afeccións e obrigas: os netos requiren avós cada vez máis versátiles e profesionais, e sabemos que tamén reclama a súa atención o magnífico viño que elabora no Ribeiro, onde é un recoñecido “colleiteiro”, seguindo a tradición familiar.

Dende a SGAPEIO, queremos deixar constancia tamén do noso aprecio e recoñecemento polo seu labor de tantos anos e desexamos a Patricio, na nova singradura que comeza, os mesmos éxitos - ou maiores - que o acompañaron nas súas tarefas anteriores.

Dirección:

Paula Raña Míguez

Comunicación coa SGAPEIO:

www.sgapeio.es
secretaria@sgapeio.es
boletin@sgapeio.es
Twitter @sgapeio
LinkedIn Sgapeio

Depósito Legal:

LU-191-1995 - I.S.S.N.:1695-7083



SGAPEIO

Sociedade Galega para a
Promoción da Estatística
e da Investigación de
Operacións

Facultade de Matemáticas
15706 Santiago de Compostela

Estatística e economía: unha relación “simbiótica”

por Patricio Sánchez Fernández

Patricio é profesor de economía na Universidade de Vigo, subdirector do Foro Económico de Galicia e socio da SGAPEIO.

Presentación

Cando se puxeron en contacto comigo dende a SGAPEIO para colaborar no seu Boletín “Informest” falando da importancia da estatística na economía, rapidamente lembrei un artigo que foi publicado conxuntamente cos meus colegas e amigos Carlos Iglesias e Esther López na Publicación Galega de Pensamento Crítico “A Trabe de Ouro” hai xa dez anos. Foi unha escusa perfecta para volver a ler ese traballo e comprobar como, malia o paso xa dunha década, o seu contido, lonxe de perder actualidade, segue estando plenamente vixente.

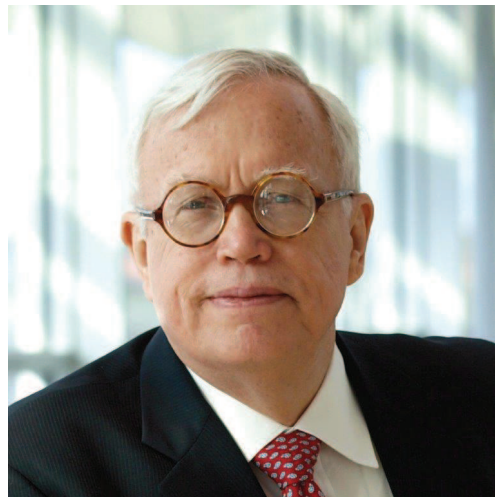
Así pois, que mellor maneira de atender á solicitude realizada que rendendo unha pequena homenaxe a ese texto e dando conta, a xeito de resumo, dos principais trazos do mesmo. Tamén desta maneira é posible expresar o meu recoñecemento e agradecemento ao traballo realizado polos meus compañeiros, así como á revista “A Trabe” por nos permitir que vise a luz unha das moitas cousas sobre as que falabamos no meu tempo como bolseiro do Instituto Galego de Estatística.

O binomio economía-estatística

A economía como ciencia é relativamente nova e, de feito, aínda subsisten eidos onde se dubida da súa consideración como tal. Non é este o obxectivo do presente artigo (aínda que daría para unha fonda reflexión) senón que, asumido o seu carácter de ciencia social, convén facer unha revisión sobre os seus principais compoñentes. Entre estes, cómpre destacar a estatística xa que supón o campo con máis rápida evolución nas últimas décadas entre os necesarios para levar a cabo unha análise económica.

Así, e consecuencia do seu crecemento exponencial no eido científico, hoxe en día, os métodos e técnicas estatísticas constitúen unha ferramenta de emprego habitual e masivo en todas as disciplinas científicas. Máis aínda, o feito cotián de que cada aspecto da actividade humana sexa medido e interpretado en termos estatísticos leva ao premio Nóbel de Economía, o profesor estadounidense James Heckman, a dicir que vivimos na “era da estatística”. Na actualidade, a compaxinación e tratamento masivo de datos posibilita un aproveitamento de todas as posibilidades ofrecidas pola estatística poñendo ao alcance da ciencia económica procedementos estatísticos de gran complexidade. Máis aínda, en nu-

merosas ocasións, moitos destes casos sería impensable telos abordado hai algúns anos cos medios dispoñibles daquela.



James Heckman, Premio Nóbel de Economía no 2000.

É por iso que a estatística, como disciplina que fai de ponte entre os modelos matemáticos e os fenómenos reais, albíscase como un elemento esencial da economía. En concreto e como noutras ciencias, pódese afirmar que o economista debe ter un coñecemento axeitado das técnicas e métodos estatísticos na realización do seu traballo. De non ser así, este non sería capaz de extraer conclusións fiables a partir dos datos dos que dispón. Asemade, o dominio destes métodos evita posibles manipulacións, ao tempo que permite participar nas argumentacións baseadas en cifras e datos.

Igualmente, dentro do amplo abano de posibilidades que ofrecen as distintas técnicas estatísticas, hai que mencionar que, ao permitir xulgar a discrepancia entre a realidade e a teoría, estas son esenciais para entender as posibilidades e limitacións da investigación económica. Do mesmo xeito, resultan fundamentais para diferenciar as conclusións obtidas a partir de datos das que, pola contra, carecen de base empírica.

Os argumentos ata aquí mencionados corroboran o estreito vencello que se dá entre a economía e a estatística. De feito, podería fálarse dunha relación recíproca e bidireccional, posto que non pode entenderse a análise económica sen unha forte base estatística, nin unha estatística sen influencia da economía, posto que as necesidades e requirimentos da mesma teñen marcado pautas na evolución dos métodos estatísticos.

Exemplos desta retroalimentación entre economía e estatística poden atoparse no caso da crítica marxiana do

sistema capitalista do século XIX que, certamente, influíu en *Bismarck* e na súa política incipiente de seguros sociais. Tamén a este respecto, cómpre mencionar o proceso de *indización* das economías, posto que o cálculo do IPC inflúe na magnitude de numerosos salarios. Aínda máis visibles se cabe, cómpre deixar constancia da influencia no desenvolvemento da estatística oficial que xorde cando determinadas magnitudes como as da contabilidade social se utilizan a nivel europeo para o reparto de fondos comunitarios ou cando as cifras de poboación se empregan para o financiamento ou na asignación do número de concelleiros que lle corresponden a un concello.

Aplicacións concretas

Co obxectivo de afondar na relación da estatística coa economía cómpre revisar as aplicacións máis importantes dos métodos estatísticos en economía. Tomando unha certa perspectiva histórica, a estatística, tal e como é coñecida dende o século XIX, xorde da confluencia de dúas disciplinas: o cálculo de probabilidades e a estatística, como termo proveniente do latín “status” e referida á ciencia do estado. A primeira vencellábase coa teoría matemática dos xogos de azar e a súa formalización pode datarse no século XVII grazas ao traballo de *Pascal* e *Fermat*, se ben é certo que as culturas primitivas xa practicaban estes xogos (tendo desenvolvidos coñecementos sobre o tema). A segunda das disciplinas, dedicada á descrición de datos, é certamente máis antiga no tempo e, así, no ano 3.050 A.C. en Exipto, xa se elaborara un censo de poboación e riqueza en relación coa construción das pirámides.

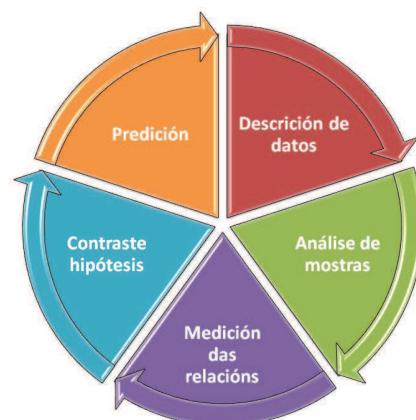
Como xa foi comentado, a evolución exponencial das técnicas estatísticas fai que na actualidade esta adquira dimensións significativamente relevantes, e semella, ao abeiro da tendencia actual, que está lonxe de rematar. Á hora de amosar as aplicacións da estatística na economía, estas pódense resumir en cinco ámbitos concretos: a descrición de datos, a análise de mostras, a medición de relacións, o contraste de hipóteses e as predicións.

A **descrición de datos**, o primeiro problema abordado historicamente pola estatística, ten como obxectivo a redución da información contida nun conxunto de datos, de xeito que teña a máxima utilidade ao ser manexada na investigación. Pola súa banda, a **análise de mostras** está fundamentada na imposibilidade de abordar (por motivos económicos ou técnicos) a toda a poboación obxecto de estudo. Así, elixindo unha mostra representativa, preténdese facer inferencia respecto a toda a poboación a partir dunha parte observada. A este respecto cómpre sinalar que esta análise ten como requisitos necesarios para a súa aplicación o contar cun modelo de probabilidade e a utilización de métodos de inferencia.

A **medición de relacións** entre as variables baséase en

que as distintas variables estudadas na análise económica non teñen unha relación exacta entre elas, senón estatística. En función disto, preténdese determinar e medir esta relación, e facer unha diagnose e validación da mesma. Seguidamente, o cuarto ámbito de relevancia consiste no **contraste de hipóteses** que ten como obxectivo poñer a proba unha determinada afirmación feita sobre unha poboación. Para isto, compáranse as conclusións resultantes da hipótese cos datos observados. Convén resaltar que esta tarefa require dunhas observacións que garantan que as conclusións obtidas estean a salvo de factores alleos.

Como aspecto de incuestionable utilidade no eido económico débese mencionar o papel xogado pola **predición**. Así, a evolución da maioría de variables amosan certa inercia de tal xeito que aínda que os seus valores futuros, como é lóxico, sexan descoñecidos, non é menos certo que a través do estudo da súa evolución histórica, pódense anticipar as pautas da evolución futura (cando menos



Ámbitos de aplicación da estatística na economía.

a curto prazo). Certamente isto proporciona un inmenso campo de oportunidades para o labor do economista.

En definitiva, as aplicacións concretas dos métodos estatísticos na economía son moitas e moi variadas. En todo caso, hai que insistir no feito de que estas cinco aplicacións non son compartimentos estancos, máis aínda, adoitan estar intimamente relacionados na práctica da economía aplicada.

Sobre este particular, débese destacar a axuda que proporcionan aos intentos de solventar os problemas motivados pola complexidade e ambigüidade dunha ciencia social como é a economía. Precisamente con esta pretensión os científicos sociais desenvolveron (e seguirán desenvolvendo) novas ramas da estatística para levar a cabo as súas investigacións. Un exemplo clásico disto vén dado pola análise factorial que, tendo a súa orixe na psicoloxía, esténdese a todas as ramas das ciencias sociais e mesmo a algunha das naturais.

Como aplicacións máis concretas basta pensar en todo o desenvolvemento do modelo lineal xeral e dos sistemas multiecuacionais ou ben na metodoloxía *Box-Jenkins* de tratamento de series temporais, mostras claras de adecuación dos métodos estatísticos aos requirimentos específicos. Asemade, cabe mencionar o desenvolvemento das técnicas de regresión semiparamétricas así como a contribución do tamén premio nóbel de economía *Robert Engle* para analizar as series temporais con volatilidade variante no tempo mediante o concepto de “heterocedasticidade autoregresiva condicional” (máis coñecido polas súas siglas en inglés “ARCH”).

Para “non” rematar: retos e esixencias

A xeito de conclusión pode resultar de interese reflectir os trazos xerais que se dan nas principais orientacións da economía e o papel que nelas poden xogar as técnicas estatísticas. Previamente cómpre sinalar que recentemente na ciencia económica, como noutros dominios e campos científicos, xeneralizáronse o emprego de categorías conceptuais, nocións e termos que implican, de feito, un novo enfoque analítico.

A este respecto, pódese dicir que existe práctica unanimidade no eido científico ao afirmar que a análise económica constitúe unha metodoloxía capaz de captar e explicar a realidade social na súa totalidade. Neste senso, mesmo resulta frecuente engadir que á economía, no estudo das relacións máis permanentes da realidade, interésanlle non soamente as uniformidades que se observan máis nitidamente en dita realidade senón tamén os fenómenos que, estando subxacentes, son igualmente definitorios.

Sen pretender formular unha relación exhaustiva, a continuación faise un percorrido sobre unha serie de cuestións que poden ser consideradas lugares comúns ao describir os elementos que se consideran esenciais para delimitar o concepto de economía. Nomeadamente, concrétnanse en catro as esixencias das análises que se levan a cabo nos distintos campos da mesma.

En primeiro lugar débese sinalar que a economía pretende substituír estudos parciais de causalidades illadas por análises xerais de interdependencia. Daquela, a preocupación fundamental da análise económica será actuar sobre unha realidade social concreta, converténdose a interdependencia xeral entre todos os fenómenos no pilar fundamental das achegas que a actividade científica ao respecto pode ofrecer. Esta interdependencia remítenos a unha certa noción de globalidade, xa que como sinala *José Antonio Alonso* na súa homenaxe ao profesor *Sampedro* (1987) “os campos aplicados da economía tratan de considerar situacións de conxunto e non análises parciais como os realizados ao nivel da microeconomía e, mesmo, da macroeconomía no senso da teoría keynesiana”.

Asemade, a economía debe dar un paso máis e estudar a realidade no seu devir histórico. Esta visión, que inclúe unha dimensión temporal, permite explicar os feitos dunha maneira eminentemente dinámica, afrontando os cambios dentro dunha totalidade dialecticamente entendida. Non se trata só de análises estáticas de interdependencia entre os elementos integrantes do campo real senón que pretende analizar a estrutura dos distintos procesos que acontecen.

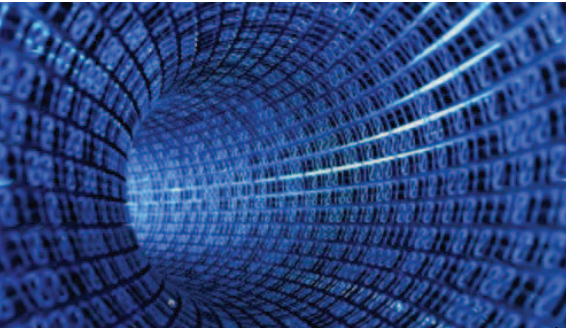


Jose Luis Sampedro, “pai” dos economistas españois.

Outro aspecto a considerar é que a economía non se debe limitar a unha mera descrición de fenómenos e relacións, permanecendo nun nivel de análise meramente descritivo e numerativo. Pola contra, debe afondar no subxacente, intentando captar, coa axuda do esforzo teórico, aquilo que non é perceptible dun modo inmediato.

As anteriores esixencias levan a concluír que a economía ten a necesidade de propiciar unha significativa apertura interdisciplinaria. Neste proceso as técnicas estatísticas teñen unha importante misión que cumprir, a cal pode ser cualificada como determinante. Igualmente hai que recalcar que esta ten que gardar sempre unha rigorosa fidelidade ao punto de partida, é dicir, á realidade específica que está a analizar. Neste senso, na análise económica desaparece toda preocupación apriorística por alcanzar universalidades ou toda predisposición a formular leis universais. Máis aínda, o que verdadeiramente importa é a inquietude por penetrar no substrato das realidades estudadas tirando á luz as estruturas a primeira vista ocultas, tarefa para a cal a estatística tórnase indispensable.

En definitiva, longa vida para a relación entre economía e estatística onde non se pode entender a súa existencia se non é dun xeito conxunto. Con toda seguridade dentro doutros dez anos, cando volvamos escribir sobre isto, serán moitos máis os exemplos que terán que ser mencionados. Só nos resta agardar...



III Xornada de Usuarios de R en Galicia

O día 20 de outubro, día Mundial da Estatística, tivo lugar na [facultade de Matemáticas da Universidade de Santiago de Compostela](#) (USC) a III Xornada de Usuarios de R en Galicia. Estivo organizada pola [Oficina de Software Libre do CIXUG](#), e patrocinada pola [AMTEGA \(Xunta de Galicia\)](#). Co obxectivo de promover e difundir o coñecemento libre da linguaxe estatística R e mostrar as súas aplicacións, leváronse a cabo relatorios durante todo o día máis dous talleres. A xornada pretende ser un punto de encontro para todas aquelas persoas interesadas en intercambiar as súas experiencias e atopar colaboracións do resto da comunidade. Pódese dicir que esta III Xornada, é como se fose a IV, xa que nos anos 2013 e 2015, celebráronse a I e II Xornada de Usuarios de R en Galicia e no ano 2014, organizáronse as VI Jornadas de usuarios de R (a nivel nacional).

O comité organizador estivo formado por Roberto Martín Souto coordinador da Oficina de Software Libre do CIXUG (consorcio das tres Universidades Galegas) e Rafael Rodríguez Gayoso, coordinador de proxectos de tecnoloxías da información no CIXUG e presidente da asociación MeLiSA (Asociación de usuarios de software libre da Terra de Melide). O comité científico estivo constituído por M^a José Ginzo Villamayor (do Departamento de Análise Matemática, Estatística e Optimización da USC) e Miguel Ángel Rodríguez Muíños (da Dirección Xeral de Saúde Pública da Consellería de Sanidade). A mesa de apertura estivo formada por M^a Victoria Otero Espinar (Decana da Facultade de Matemáticas - USC), Mar Castro Pereiro (Coordinadora da OCFLOSS - AMTEGA), Salvador Naya Fernández (Vicerreitor de Política Científica, Investigación e Transferencia - UDC), Miguel Ángel Rodríguez Muíños (Representante de MeLiSA) e Roberto Martín Souto (Coordinador da Oficina de Software Libre - CIXUG).

O programa da Xornada estivo dividido en 4 bloques. No primeiro, Roberto Martín (coordinador da Oficina de Software Libre do CIXUG), contou toda a labor que se leva a cabo dende a OSL. O título da súa comunicación foi “A OSL con R”. No segundo bloque, contouse con relatores da administración pública:

- “*Pyramid_plot: unha nova librería para a representación de pirámides de poboación*” impartida por Marcos Fernández da AMTEGA.
- “*Elaboración da clasificación do grao de urbanización das parroquias galegas (GU2016) empregando R*” impartida por M^a Esther López Vizcaíno do Instituto Galego de Estatística (IGE). M^a Esther Calvo Ocampo, María Martín Vila, Estefanía Villar Cheda, Carlos Iglesias Patiño e Solmary Silveira Calviño son coautores deste traballo. Presen-

touse o procedemento utilizado para obter a clasificación do grao e do subgrao de urbanización das parroquias galegas (GU 2016) empregando a metodoloxía establecida por Eurostat no ano 2011.

- “*R-Shiny: unha ferramenta para mellorar a difusión das operacións do Sistema de Contas Económicas de Galicia*” impartida por Noa Veiguela Fernández do IGE. Contáronse mecanismos empregados no IGE para a difusión da información estatística e exemplos de aplicacións web creadas con R-Shiny.
- “*Automatización de procesos con R. Importación de datos de saúde. Un caso práctico: A gripe en Galicia*” impartida por Miguel A. Rodríguez Muíños da Consellería de Sanidade. M^a Jesús Purriños Hermida e Gael Naveira Barbeito son coautoras deste traballo. Presentouse o método empregado no SERGAS para predicir con R a onda da gripe, así como o cálculo dos límites da mesma.

No terceiro bloque, contouse con profesores das tres universidades galegas:

- “*Control estadístico da calidade mediante las librerías qcr, ILS y fda.qc*” impartida por Javier Tarrío Saavedra e Salvador Naya Fernández da Universidade da Coruña. Miguel Flores é coautor deste traballo. Mostráronse e explicaron as utilidades para o control estatístico da calidade proporcionados polos novos paquetes de R *qcr* (quality and reliability review), *ILS* (interlaboratory study) e *fda.qc* (functional data analysis quality control).
- “*Plugin RCommander para o ensino da Inferencia Estatística na Universidade de Vigo*” impartida por Manuel Alfredo Mosquera Rodríguez da Universidade de Vigo (UVigo). Tomás R. Cotos Yáñez, Ana Pérez González e Benigno Reguengo Lareo son coautores deste traballo. Presentouse o *plug-in* de Estatística Básica con R-Commander. Este *plug-in* serve de apoio na docencia sobre todo para aqueles alumnos que non están familiarizados coa programación.
- “*Serie de tempo con R*” impartida por Manuel Febrero Bande da USC. Fíxose un amplo repaso pola teoría de series de tempo empregando R con diferentes exemplos.

O último bloque correu a cargo dun alumno do Máster en Técnicas Estatísticas (Marcos Matabuena Rodríguez) e 9 egresados do mesmo Máster, dos cales 8 son alumnos de Doutoramento e un Doutor (Manuel Arias Rodil) dende novembro de 2015, todos eles da USC.

- “*Paquete VGAM. Aplicaciones biomédicas con R*” impartida por Jenifer Espasandín Domínguez. Carmen M^a Cadarso Suárez, Thomas Kneib e Francisco Gude son coautores deste traballo. No traballo preséntanse modelos GAMs vectoriais (VGAM, Vector Generalized Linear and Additive Models, Yee and Wild, 1996). Utilizando o paquete VGAM de R (Yee, 2015) ilústrase a utilidade destes modelos con datos reais biomédicos.
- “*InventorymodelPackage: os problemas de inventario multi-axente*” impartida por Alejandro Saavedra Nieves. O paquete *InventorymodelPackage* de R completa o estudo teórico dos principais modelos de inventario multi-axente introducidos na literatura.
- “*Una introducción a la teoría de colas desde R*” impartida por Marcos Matabuena Rodríguez. Presentouse unha breve introdución á teoría de colas, a través dunha primeira introdución conceptual, explicando os elementos básicos que caracterizan unha cola e posteriormente ilustrar os modelos de colas máis sinxelos, desde un punto de vista práctico, a través do software estatístico R e o paquete *queueing* creado por Pedro Canadilla.
- “*Tratamento de modelos paramétricos de terremotos con R*” impartida por María Isabel Borrajo García. Este traballo consiste na implementación en R dunha nova variante do modelo *Epidemic Type Aftershock Sequence* (ETAS), o uso do mesmo para xerar unha serie de catálogos de terremotos e a realización dunha análise de sensibilidade sobre os parámetros involucrados e de pares dos mesmos, na función de intensidade.
- “*Creating a Board Game with R*” impartida por Jose Ameijeiras Alonso. O obxectivo desta charla foi mostrar como se pode crear un xogo de mesa co software R. Para iso, escolleuse o popular xogo de mesa de bens raíces cuxo obxectivo é o de conseguir un monopolio, posuíndo todas as propiedades inmobles que aparecen no xogo.
- “*Regresión cuantil lineal local. Un nuevo paquete de R*” impartida por Mercedes Conde Amboage. Este traballo céntrase no marco da regresión cuantil, na selección do coñecido parámetro xanela no contexto da estimación lineal local tipo núcleo. Preséntase un paquete en elaboración de R que conterá diversas funcións que permitan computar os diferentes selectores do parámetro xanela dispoñibles na literatura, así como unha nova proposta de selector *plug-in*.
- “*Mapas con R*” impartida por María José Ginzo Villamayor. Neste traballo móstranse 4 formas diferentes de crear mapas con R, empregando distintos paquetes.
- “*R como ferramenta para a análise de datos xenómicos*” impartida por Pilar Cacheiro Martínez. No eido da xenómica, as tecnoloxías de xenotipado e ultrasecuenciación están a xerar cantidades inxentes de datos. Os datos producidos, logo dun procesado inicial que normalmente se fai empregando programas específicos, poden a continuación ser manexados e analizados con R.
- “*galfoR: paquete de R para a xestión forestal en Galicia*” impartida por Manuel Arias Rodil. José Mario González González e Ulises Diéguez Aranda son coautores deste traballo. Neste traballo preséntase unha proposta de estrutura para o paquete, nomeado *galfoR*, que incluírá os modelos dinámicos e o conxunto de algoritmos para simular o crecemento das principais especies forestais de Galicia. O paquete empregará como *inputs* dous obxectos: unha lista coa información que caracteriza o rodal, e outra lista coas variables que definen a alternativa selvícola.

Para rematar houbo dous talleres, un para aqueles que se inician con R e o outro para usuarios máis avanzados: “*Taller práctico e básico de R*” impartido por M^a José Ginzo Villamayor da USC e o “*Taller Reproducibilidad en ciencia con R. Da minaría de datos á publicación*” impartido por Miguel Blanco Vázquez da UVigo. Os talleres ademais de actividades da xornada enmárcanse dentro das actividades da Aula de Matemáticas da Facultade de Matemáticas, coa que colabora estreitamente a CI-XUG.

Pódese dicir que a Xornada acadou unha alta participación, con máis de 70 persoas inscritas. Así mesmo, os talleres, de case 2 horas de duración, contaron con moitos asistentes, arredor de 25 persoas en cada un deles.

Quero rematar agradecendo, no nome do Comité Científico, a colaboración de todas as persoas (poñentes, moderadores, asistentes á xornada, seguidores *online*, *twiteiros*...) que fixeron posible esta III Xornada, e as contribucións dos patrocinadores e colaboradores. Grazas tamén á OSL pola iniciativa de organizala.



Información Adicional

Para máis información sobre a III Xornada de Usuarios de R en Galicia véxase a páxina web: <https://www.r-users.gal/>, onde se poderán atopar as presentacións e vídeos da xornada.



Actividades da SGAPEIO

Sesión da SGAPEIO no XXXVI Congreso Nacional de Estadística e Investigación Operativa

Esther López Vizcaíno

Con motivo da celebración do XXXVI Congreso Nacional de Estadística e Investigación Operativa (Toledo, 5-7 de setembro de 2016), organizado pola Sociedad Española de Estadística e Investigación Operativa (SEIO), a SEIO invitou á SGAPEIO a organizar unha sesión específica da nosa sociedade no dito congreso, co obxectivo de potenciar os lazos xa existentes entre as dúas sociedades e dar a coñecer aqueles temas de traballo que están nas fronteiras das nosas áreas de interese.



Os temas de traballo da SEIO e da SGAPEIO son semellantes, a única diferenza pode vir pola aplicación. Dende a SGAPEIO pensamos que sería oportuno organizar unha sesión específica onde a característica fundamental fose que as presentacións versasen sobre métodos e aplicacións da Estatística e a Investigación Operativa en Galicia, é dicir, que os traballos ao final se aplicasen a datos que tivesen que ver con Galicia.

Unha vez decididas as características que deberían cumprir as presentacións, fíxose un chamamento entre todos os socios da SGAPEIO co obxectivo de que participasen nesta sesión invitada pola SEIO. Finalmente presentáronse as seguintes ponencias cos títulos:

- Xestión de recursos na extinción de incendios forestais en Galicia: Un modelo para o control dos recursos aéreos (J. Rodríguez Veiga, I. Gómez Costa, M. J. Ginzo Villamayor, B. Casas Méndez, J. L. Saiz Díaz).
- Análise de supervivencia das empresas de Galicia: efecto de endeudamento e rendibilidade. (A. Vaa-

monde Liste, M. Meijide Vecino, P. Sánchez Bello, R. Luaces Pazos, P. Sánchez Fernández).

- Aproximación á carga de poboación estacional dos concellos de Galicia (P. Romero Martínez, E. López Vizcaíno).
- O Criterio de Información de Akaike para a estimación dos ocupados por actividade económica (E. López Vizcaíno, M. J. Lombardía Cortiña, C. Rueda Sabater).

A sesión celebrouse o martes día 6 e estivo moderada por min. Comenzou coa exposición da primeira ponencia por parte de M^a José Ginzo, que nos falou sobre a xestión de recursos na extinción de incendios forestais en Galicia. Neste traballo presentaron técnicas de programación lineal enteira co obxectivo de servir de apoio aos coordinadores das tarefas de extinción de incendios de forma rápida. A continuación, Antonio Vaamonde fixo unha exposición sobre a supervivencia das empresas de Galicia. Para a análise de supervivencia aplicaron diferentes técnicas estatísticas como o método de *Kaplan-Meier*, regresión de Cox, ou diversos contrastes para valorar a significación dos factores que inflúen na vida e desaparición das empresas. Despois da supervivencia de empresas, Pilar Romero falounos de como poder aproximarse á carga de poboación estacional que soportan os concellos galegos. Para analizar a carga real de poboación que soportan os concellos galegos é necesario coñecer os movementos que se producen no ano por motivos de traballo, estudos, turismo, saúde... Como a información de que se dispón non é exhaustiva, é necesario recorrer a modelos estatísticos como os loxísticos e os de áreas pequenas, que nos permitan aproximarnos ás magnitudes desexadas. E xa por último, eu mesma, contei como estimar a poboación ocupada en Galicia por actividades económicas (agricultura, pesca, silvicultura,...) utilizando técnicas de estimación en áreas pequenas e empregando o Criterio de Información de *Akaike*.

A sesión resultou moi enriquecedora, con moita audiencia e interesantes discusións, co cal no nome da SGAPEIO quero agradecer a todos os ponentes a dispoñibilidade para participar nesta sesión específica achegándonos os seus traballos. Por outra parte, tamén queremos agradecerlle á SEIO a súa invitación para formar parte das sesións deste XXXVI Congreso. A experiencia foi moi positiva e enriquecedora e espero que se poida seguir mantendo este tipo de colaboracións no futuro.



Matriz de contabilidade social

A Matriz de Contabilidade Social (MCS) é unha operación estatística que debe contextualizarse no Sistema de Contas Económicas de Galicia cuxo núcleo principal está formado polas Contas económicas de Galicia e o Marco *Input-Output* de Galicia, ambas as operacións consolidadas segundo o Plan galego de estatística. O principal obxectivo desta operación de síntese é servir de vínculo entre os procesos de xeración de valor engadido e a descrición dos fluxos de bens e servizos dunha economía, recollidos nunha Táboa *Input-Output* (TIO), e o papel das persoas e os fogares na economía. Proporciona, polo tanto, información de carácter social, pero o que a distingue doutras estatísticas máis específicas é que ofrece consistencia entre as estatísticas sociais e as contas económicas, isto é, ofrécenos unha visión combinada das cifras macroeconómicas e a súa repercusión nas persoas da nosa economía. Segundo a MCS entre o ano 2008 e o 2013 a masa salarial en Galicia descendeu un 12 %. Por sexos esta caída concéntrase nos homes cuxa masa salarial diminúe nun 17 %, mentres que nas mulleres o fai nun 4 %. Consecuentemente, os soldos e salarios da poboación masculina perden peso: no ano 2008 supoñen o 62 % da masa salarial e o 58 % no 2013. Por outra parte, no ano 2013, aproximadamente o 53 % dos fogares ten como principal fonte de ingresos as rendas procedentes do traballo (soldos e salarios e rendas mixtas). O 47 % restante repártese entre os fogares nos que as pensións contributivas supoñen a principal fonte de renda (38 %) e os que dependen das prestacións por desemprego, pensións non contributivas e outras rendas.

Pódese obter máis información en: [web](#)

Explotación da mostra continua de vidas laborais

A Mostra continua de vidas laborais (MCVL) é un conxunto de microdatos anonimizados procedentes de diversos rexistros administrativos: da Seguridade Social, do Padrón municipal de habitantes e da Axencia Tributaria. A MCVL constitúe unha mostra representativa de todas as persoas que durante o ano tiveron relación coa Seguridade Social, ben porque estiveron afiliadas en situación de alta laboral, ben porque percibiron unha prestación contributiva ou subsidio por desemprego, ou ben porque percibiron algún tipo de pensión contributiva da Seguridade Social. Durante o ano 2014, 1.092.075 galegos en idade laboral (de 16 ou máis anos) mantiveron algunha relación de afiliación en alta laboral coa Seguridade Social, un 0,1 % máis que no ano 2013. O 3,4 % destes afiliados traballou menos dun mes en todo o ano 2014, o 13,1 % entre un mes e seis meses, o 17,6 % traballou seis ou máis meses pero sen chegar ao ano completo e, finalmente, o 65,9 % restante fíxoo durante todo 2014. Por outro lado, o 24,3 % dos galegos que en 2014 estiveron

afiliados en alta laboral por conta allea percibiu un salario total menor ou igual aos 6.000 euros anuais. Entre os traballadores que só estiveron afiliados en alta laboral parte do ano, esta porcentaxe sobe ata o 54,6 %, mentres que para os que estiveron afiliados o ano completo é do 4,1 %. Para este último colectivo, o intervalo modal (o de maior frecuencia) é o de 15.000 a 20.000 euros, no que se atopa o 26,2 % dos traballadores que estiveron afiliados en alta laboral durante todo o ano 2014.

Pódese obter máis información en: [web](#)

Plan galego de estatística 2017-2021

O 2 de agosto de 2016 publicouse no Diario Oficial de Galicia (DOG Núm. 145) a Lei 15/2016, do 28 de xullo, do Plan galego de estatística 2017-2021. Xunto coas 101 metas xenéricas de información que se van satisfacer nos programas anuais, o plan inclúe 117 estatísticas concretas que se conseguiron consolidar nos plans precedentes e que se deberán executar. A maioría destas estatísticas concretas son novas con respecto ao plan anterior, e abordan temas como as empresas TIC de Galicia, a poboación dependente, o desprazamento transnacional de persoas traballadoras, a estatística de emerxencias 112 ou a ocupación na pesca. O novo plan estatístico contempla obxectivos informativos, instrumentais e de calidade, así como a obriga de considerar a perspectiva de xénero na información estatística. Tamén se reforzan as actividades de análise para aumentar o valor da información dispoñible e a satisfacción das persoas e entidades usuarias. O plan asume ademais o código de conduta adoptado polo comité do Sistema Estatístico Europeo. Polo tanto, a organización estatística da Comunidade Autónoma de Galicia queda sometida aos seus principios e o Goberno autonómico a impulsalos e respectalos. A independencia profesional do persoal estatístico, a imparcialidade, a obxectividade e a transparencia son algúns deles. Máis información en: [web](#)

Consello Reitor do IGE

O pasado 13 de setembro constituíuse o Consello Reitor do IGE, previsto no seu Estatuto como órgano superior colexiado de goberno. Presidido polo Conselleiro de Facenda e con representación dos departamentos competentes en persoal, avaliación, orzamentos e administracións públicas, a súa creación responde á necesidade de ordenar un conxunto de tarefas que xa se viñan realizando para dar soporte ás funcións estatísticas do IGE. En particular, o Consello Reitor aprobará o anteproxecto de orzamentos e a proposta de relación de postos de traballo do Instituto.

Oposicións

O 26 de setembro de 2016 publicouse no Diario Oficial de Galicia (DOG Núm. 183) as listaxes provisionais de persoas admitidas e excluídas na Escala Superior de Estatísticos. Máis información en: [web](#)

Traballos de Estatística e Investigación Operativa no SUG

Teses

Contributions to Nonparametric Estimation of Survival Curves in Non-Markov Multi-State Models

Leyla Azarang.

Director: Jacobo de Uña Álvarez.

Data da defensa: 27 de outubro de 2016. UVigo.

Esta tese resolve distintos problemas metodolóxicos da análise estatística de datos de supervivencia, e aplica os métodos propostos ao estudo da enfermidade do Lupus en España.

En concreto, na tese demostrase a consistencia do estimador *Jackknife* da covarianza entre dúas integrais de *Kaplan-Meier* multivariantes, resultado que se aplica inmediatamente á estimación da varianza da matriz de probabilidades de transición empíricas dun proceso multi-estado non markoviano, moi relevante na análise de datos biomédicos. Estes resultados veñen a completar o estudo das integrais de *Kaplan-Meier* con covariables iniciado nos anos noventa polo profesor Winfried Stute.

Ademais, nesta tese propónse unha metodoloxía estatística flexible para a detección de factores de risco en modelos multi-estado, mediante a estimación do impacto que teñen tales factores sobre as probabilidades de transición. Para facilitar a aplicación práctica de tales métodos, deseñouse unha librería de *software* libre no [entorno R](#).

A tese de doutoramento conclúe co estudo de 3630 casos de pacientes diagnosticados de Lupus Eritematoso Sistémico en España, recollidos no rexistro RELESSER. Este rexistro, elaborado pola Sociedade Española de Reumatoloxía, permitiu nos últimos anos o avance do coñecemento sobre a enfermidade do Lupus en España. Os novos resultados aportados por esta tese permiten comparar ao longo do tempo a supervivencia libre de dano entre homes e mulleres, así como para pacientes de distinta idade, ao diagnóstico. Unha das conclusións que arroxa o estudo refire a maior supervivencia libre de dano que teñen as mulleres fronte aos homes. Por outra parte, comprobouse que o pronóstico sobre a evolución da enfermidade empeora de maneira progresiva coa idade, sendo máis favorable para pacientes diagnosticados en idades temperás. Os métodos levados a cabo na tese poderían ser utilizados no futuro para personalizar aínda máis esta investigación, considerando as distintas tipoloxías de dano que poden presentar os pacientes.

Traballos fin de Grao en Matemáticas na área de Estatística

Os traballos presentados nesta sección foron defendidos ao longo do curso 2015–2016 na facultade de Matemáticas da USC como Traballos Fin de Grao dirixidos dende o Departamento de Estatística e Investigación de Operacións.

Splines como instrumento de aproximación e estimación

Álvaro Lence Burgos.

Director: César Sanchez Sellero.

Neste traballo preséntase unha revisión do método dos *splines* como forma para aproximar e estimar a recta de regresión. Para isto descríbense os distintos tipos de *splines*, e inclúense tamén varios procesos para facilitar o seu cálculo. Posteriormente expónse unha maneira de

elixir o parámetro de suavización, a validación cruzada, e preséntase un método para estimar a varianza residual. No último capítulo ofrecemos os resultados dun estudo que levamos a cabo con datos reais e datos simulados. Neste estudo aplicáronse os métodos *splines* expostos e comprobáronse as súas propiedades estatísticas como estimadores da función de regresión. En concreto, analizouse o sesgo, a varianza e o erro cuadrático medio, tanto puntual como integrado, como medidas da calidade do estimador, e estudouse a súa relación co parámetro de suavización, tanto en valores fixos do mesmo como para o selector por validación cruzada xeneralizada.

Unha introdución á remostraxe de modelos de regresión lineal

Daniel Barreiro Ures.

Director: José Manuel Prada Sánchez.

A remostraxe confórmana todas aquelas técnicas estatísticas que apliquen un certo procedemento estatístico sobre unha mostra observada co fin de obter dela a máxima información posible e levar a cabo a inferencia. A gran diferenza con respecto á teoría de inferencia clásica ou normal radica no feito de que esta require dunha serie de hipóteses (normalidade, homocedasticidade, independencia, etc.) acerca da poboación á que pertencen os datos e o mecanismo que os xera. Pola súa parte, as técnicas de remostraxe non impoñen a maioría desas hipóteses aínda que, para o seu correcto funcionamento, é crucial que a mostra observada sexa representativa da poboación e, nalgúns casos, que os datos fosen obtidos independentemente uns doutros.

O mecanismo que xere os datos pode ser, no caso máis simple, unha función de distribución, ou algo máis complexo como un modelo de regresión, caso de especial interese neste traballo.

O traballo versará sobre tres técnicas de remostraxe: o test de permutacións, a validación cruzada e, especialmente, o *bootstrap*. Ditas técnicas aplicaranse en exemplos de simulación, comparando o seu comportamento co da inferencia clásica, e, finalmente, sobre dous conxuntos de datos reais.

Inferencia sobre cuantís

Daniel García Rasines.

Directores: César Sánchez Sellero, Mercedes Conde Amboage.

Os cuantís veñen acadando cada día unha maior atención na Estatística, como consecuencia da súa capacidade para describir de maneira máis completa a distribución dunha variable tanto nas súas posicións centrais como nas súas posicións inferiores ou superiores. Ademais, mentres que a media é moi sensible a datos atípicos, os cuantís son medidas robustas que se adaptan a toda clase de situacións, sen necesidade de supoñer que os datos proveñen dunha distribución normal. Este traballo ten por obxectivo presentar de maneira detallada distintas técnicas empregadas na inferencia non paramétrica sobre cuantís. Na primeira parte introdúcese o concepto de cuantil e estúdanse as propiedades máis relevantes do seu estimador máis estendido, o cuantil mostrál. Seguidamente, preséntanse diversos estimadores do erro típico dos cuantís mostrais e realízase un estudo de simulación para comparalos mediante o *software* estatístico R. Os

dous últimos capítulos do traballo dedícanse, respectivamente, á construción de intervalos de confianza para cuantís e á descrición de *tests* que permiten contrastar hipóteses referidas a estes parámetros. Coa finalidade de facilitar o seguimento dos procedementos teóricos, inclúese un apéndice coas definicións e resultados de teoría da probabilidade que se empregan no desenvolvemento dos mesmos. Ademais, inclúese un segundo apéndice que contén o código de R correspondente ao estudo de simulación.

Modelos de investigación operativa en problemas de xestión de inventarios

Teresa Veiga Rodríguez.

Directora: Balbina V. Casas Méndez.

O contexto socioeconómico actual, a globalización e o aumento incesante da demanda fixeron da loxística un dos aspectos máis importantes para todas aquelas empresas que se dedican a ofertar produtos. Hoxe máis ca nunca, a xestión de inventarios supón unha vantaxe competitiva. Non obstante, o uso dos inventarios non é algo recente, estes nacen na antigüidade como un sistema para evitar a escaseza. Máis adiante, en plena revolución industrial, propóñense os inventarios como unha ferramenta para lograr unha competitividade empresarial, xurdindo os modelos de xestión de inventarios como resposta científica ao proceso de conseguir unha política óptima de funcionamento. Dende entón ata hoxe en día son numerosos os avances que houbo, sendo un campo de estudo moi de actualidade e en constante renovación. Non é posible comprender os modelos actuais sen unha base de modelos de inventarios. E é esa introdución aos modelos de xestión de inventarios no que se centra este traballo. Primeiramente, motívase a existencia dos modelos de inventarios e detállanse as variables involucradas para, ao longo do documento, estudar diferentes modelos. Comézase cos modelos máis básicos, modelos deterministas, para chegar a modelos máis complexos e realistas, como os modelos de múltiples chanzos ou os probabilísticos. Por último, introdúcense brevemente os xogos de inventario e contéplanse ferramentas informáticas que dan solución aos modelos propostos.

Implementación do método das submatrices

Branca García Correa.

Director: José Manuel Prada Sánchez.

A memoria comeza cunha breve introdución á teoría de xogos, comezando pola chamada teoría da decisión e

continuando cos xogos en forma estratéxica. Dentro deles, profúndase nos xogos bipersoais de suma nula, nas estratexias mixtas en xogos finitos e remátase coa introdución dos xogos matriciais. Todo iso ilústrase con exemplos e demóstranse todos os resultados utilizados. O traballo continúa coa explicación dun método para resolver xogos matriciais de calquera dimensión, o método das submatrices, baseado na caracterización dos puntos extremos das estratexias óptimas dos xogadores. En primeiro lugar, vanse enunciar e probar os tres teoremas que supoñen o sustento teórico para o devandito método: o teorema de Krein-Milman e dous teoremas debidos a Shapley-Snow. A continuación realízase a implementación do método das submatrices mediante o software estatístico R. Para finalizar, explícase por que o método das submatrices é un método máis completo que a programación lineal á hora de resolver xogos matriciais, posto que é capaz de dar todas as solucións óptimas simultaneamente. Tamén se ilustra este feito cun exemplo.

O algoritmo de Barrodale e Roberts

Santiago Díaz Candán.
Director: César A. Sánchez Sellero.

Na estatística frecuentemente recórrase ao uso da regresión en media, pero existen outras posibilidades. A regresión cuantil diferénciase da media en que a estimación se produce minimizando unha perda cuantílica no canto duns mínimos cadrados, o cal pode resultar moi útil en determinados contextos. O funcionamento da mínima perda cuantílica consiste na minimización dun sistema de ecuacións lineais, e aquí é onde entra en acción o algoritmo de Barrodale e Roberts. Polo tanto, o traballo estruturárase en base a tres eixos:

- Unha breve introdución á regresión estatística, dotando a esta sección das definicións necesarias para a comprensión dos diversos tipos de regresión, entre os que destacaremos a de tipo media, mediana e cuantil.
- Explicación do funcionamento do método simplex, algoritmo que permite a resolución de sistemas lineais, e cuxa optimización constitúe o algoritmo de Barrodale e Roberts.
- Tras incluír unhas nocións básicas sobre o problema de programación linear e a aproximación l_1 , estableceranse exemplos introdutorios ao algoritmo de Barrodale e Roberts que nos permiten ver o seu funcionamento, para, a continuación, defini-lo formalmente adaptándoo ao caso da regresión cuantil.

Técnicas de aprendizaxe estadística no contexto de "Big Data"

Laura Freijeiro González.
Director: Wenceslao González Manteiga.

Este traballo analiza os modelos de regresión e as regras de clasificación na análise discriminante. Mentres que os modelos de regresión permiten obter unha predición do valor dunha variable, en base a uns novos valores das variables explicativas fixados. As regras de clasificación da análise discriminante posibilitan clasificar unha observación futura nun determinado grupo dentro dos k posibles aos que podería pertencer. Descríbense ambos os campos, cunha motivación deles e os diversos métodos empregados para levar a cabo os fins propostos en cada un. En todos os procedementos ponse de manifesto a situación da alta dimensión no número de variables predictoras na regresión e no número de variables clasificadoras na análise discriminante, que busca respostas e método para manexar un gran número de datos en campos tan diversos e importantes como poden ser o da medicina ou as finanzas, entre outros. Propóñense os problemas que presenta e procedementos para resolvelos. Finalmente, vaise ilustrar a teoría exposta cun exemplo a través do código implementado en R.

Regresión Segmentada

Manuel Antonio Novo Pérez.
Directora: Rosa M. Crujeiras Casais.

En ocasións, nun modelo de regresión, a relación entre a variable explicativa e a variable resposta cambia en distintos intervalos da variable explicativa. Neste traballo tratamos un destes casos: o modelo de regresión segmentada, que vén ser un modelo de regresión no cal a función de regresión é una liña recta en cada un dos intervalos nos que segmentamos a variable explicativa. Comezamos introducindo o concepto de regresión e vendo algúns modelos famosos. A continuación, introducimos que é a regresión segmentada e formulamos o modelo correspondente. Despois vemos como estimar e facer inferencia sobre os parámetros do modelo cando o punto de ruptura é fixo, comentando tamén o modo de facer predicións utilizando o modelo e como validar este último. Logo vemos un método para estimar e facer inferencia sobre os parámetros do modelo, con puntos de ruptura descoñecidos. Tratamos tamén como detectar a localización e o número dos puntos de ruptura e como facer predicións a partir do modelo. Rematamos realizando unha simulación para comprobar o funcionamento dos estimadores dos parámetros do modelo.

Técnicas de aprendizaxe estatística. Clasificación mediante regresión loxística

Sabela Fernández Villar.
Directora: Beatriz Pateiro López.

O auxe da información e dos datos que se está a producir na sociedade actual require de métodos capaces de interpretar e analizar ditos datos masivos (fenómeno coñecido como *Big Data*). Isto é precisamente o que a aprendizaxe automática proporciona. Introdúcese o concepto de aprendizaxe automática e faise unha breve descrición dun dos métodos máis relevantes da aprendizaxe automática: o método de clasificación. En concreto, un dos métodos habituais para tratar problemas de clasificación é a regresión loxística. Realízase unha revisión exhaustiva do modelo loxístico binario así como unha ampliación do mesmo. Desta forma, utilizando o modelo loxístico binario, vense aproximacións para predicir respostas cualitativas binarias, mediante o proceso de clasificación. Detállase o procedemento xeral da máxima verosimilitude para o axuste dos parámetros do modelo loxístico co fin de realizar inferencia estatística sobre el. Descríbense as técnicas de selección de variables e de re-

gularización, que son fundamentais no contexto do *Big Data*. E, por último, realízase unha breve introdución ao modelo loxístico politémico, como ampliación do modelo loxístico binario.

Modelización matemática en problemas de loxística usando AMPL

Lucía Tajuelo López.
Directora: Balbina Casas Méndez.

No sector da loxística o gasto destinado ao transporte representa un dos maiores gastos para as empresas, así que, poder optimizar estes procesos de forma axeitada é un punto clave neste ámbito. O obxectivo principal deste TFG é o estudo de dúas cooperativas agrarias que precisan mellorar as rutas dos seus camións. Para isto precisa- rase enmarcar teóricamente os problemas de enrutamento, cos que modelaremos os problemas das cooperativas, e presentar unha linguaxe de programación axeitada, neste caso utilizamos AMPL, para verificar os modelos creados. Finalmente refináronse os modelos creados co uso de métodos heurísticos.

Traballos Fin de Máster

Traballos do Máster Interuniversitario en Técnicas Estatísticas presentados en setembro de 2016.

Modelos de regresión con alta dimensión no número de covariables

Iván Dopazo Iglesias.
Director: Wenceslao González Manteiga.
USC, modalidade académica.

Na actualidade, dispoñemos dunha gran cantidade de datos, tanto no ámbito da economía, como da bioloxía, xenómica, etc. A xeneralización do uso de páxinas web, redes sociais, dispositivos móbiles, etc, fai que a cantidade de datos recollidos e almacenados non deixe de crecer cada día, polo que se está facendo cada vez máis complexa a tarefa de analízalos. Á hora de analizar estas grandes cantidades de datos, os métodos estatísticos que se utilizaban deixan de ser tan eficientes, polo que é necesario desenvolver novos métodos. O obxectivo deste traballo é expoñer algúns métodos que nos permitan realizar a análise destas grandes cantidades de datos, centrándonos no caso de que o número de variables que incorporemos ao modelo (p) sexa maior que o número de casos (n), tanto na situación en que a variable resposta dos modelos sexa continua como binaria.

Modelos de predición clínica

Carla Díaz Louzao.
Directora: Carmen María Cadarso Suárez.
Titor: Francisco Gude Sampedro (CHUS).
USC, modalidade prácticas.

Os modelos de diagnóstico e prognóstico son moi importantes en moitos eidos, incluíndo a medicina. Este tipo de modelos, que estiman de forma individualizada o risco de morte ou de presentar unha enfermidade para os pacientes, estanse volvendo cada vez máis importantes na investigación realizada en hospitais debido aos seus posibles usos, como son distinguir estadios dunha enfermidade e clasificar a súa severidade ou avaliar a supervivencia tras un procedemento ou a diagnose dunha enfermidade, xeralmente co obxectivo de tomar decisións diagnósticas ou de tratamento.

Neste traballo trátanse dous tipos de análise: unha primeira, centrada no diagnóstico dunha enfermidade en función doutras variables e unha segunda, consistente nunha análise de supervivencia. A principal motivación detrás deste traballo é a aplicación destes modelos a dous casos prácticos, un relativo ao diagnóstico de cancro e o outro relativo á supervivencia de enfermos cardíacos despois de seren sometidos a terapia de resincronización cardíaca.

Coñecéndonos

Coñecendo a un socio da SGAPEIO: Miguel Ángel Rodríguez-Gigirey Pérez.

Son Miguel Ángel Rodríguez-Gigirey Pérez, natural de Ourense aínda que por cuestión de traballo vivín nun montón de vilas nas catro provincias galegas. Estudei matemáticas en Santiago e levo 24 anos de profesor de ensinanza secundaria. Actualmente vivo en Santiago de Compostela e traballo no instituto de Arzúa.



Cantos anos levas traballando en estatística ou IO?

Non especificamente en estatística, senón en matemáticas en xeral, dando clase, pero sempre me pareceu moi interesante esta parte nas miñas clases e tratei de fomentala.

Descríbenos o teu traballo

É fácil e coñecido por todos, profesor de ensinanza secundaria, na especialidade de matemáticas.

Dende cando es socio/a da SGAPEIO?

Desde o primeiro ano. Non fun socio fundador, pero tan pronto tiven noticia dela, fíxenme socio. Eu diría que sobre o ano 92 ou 93, penso que foi por aí cando se fundou.

Cal foi o 1º congreso da SGAPEIO ao que asistiches?

Pois estou convencido de que xa estiveron no I Congreso, que fora en Coruña, pero o primeiro que lembro ben foi o II Congreso en Vigo.

O tema do teu próximo proxecto/artigo?

Actualmente non teño tempo para nada específico no tema, ando algo liado con outras cousas de traballo e

non me permiten actividades extras.

Un curso que che gustaría facer?

Un curso de R. Estou aprendendo algo coa miña filla, que estuda matemáticas e o usa. Paréceme interesante. Xa sei que hai tempo que se utiliza, pero como digo, ando algo liado agora.

Que libro de estatística ou I.O. recomendarías?

Pois utilizo sempre, cando teño dúbidas de algo, o libro *Introducción a la estadística y sus aplicaciones*, de Ricardo Cao e outros. Creo que todos os autores son profes da UDC, algúns deles foron compañeiros na directiva da SGAPEIO cando estiven eu.

Se non foses estatístico/a, que che gustaría ser?

Creo que se non fose matemático non serviría para nada máis!!!

Se oes “estatística”, que é o primeiro que che vén á cabeza?

Os titulares dos xornais do tipo “Un de cada tres galegos...” e cousas dese tipo. Sempre me gusta mirar nas tripas deses artigos e mirar se teñen base científica, estudos ben feitos, ou simplemente titulares con pouco rigor.

Que é o que máis valoras do labor da SGAPEIO?

Recoñezo que ultimamente estou moi desconectado da sociedade, por falta de tempo. Sempre me pareceu importante darlle difusión a esta parte das matemáticas, ás veces pouco valorada entre os propios profes matemáticos. Cando estaba máis pendente, sempre me gustaron os cursos e os congresos.

En que cres que deberíamos mellorar?

Sempre me pareceu que se ocupaba pouco na parte de secundaria, incluso cando o “responsable” desta parte era eu. Non sei ben onde está a solución, nin como fomentala. Paréceme que o que se fai está ben, pero o certo é que non chega nin atrae a este colectivo, penso. Creo que tamén, nalgún tempo, se fixeron máis cursos e charlas dedicados a secundaria e á estatística base e quizais agora os congresos son a máis alto nivel, e algúns perdémonos. Tampouco sei onde está a solución nin como enganchar a máis xente.



Sabías que?

Poñamos as cartas enriba da mesa por Gael Naveira Barbeito

Toda boa partida de cartas debe comezar por un bo "barallado" das mesmas, pero, que quere dicir esta expresión? Esta expresión significa que as cartas deben ser mesturadas ata que cada unha delas quede totalmente aleatorizada no mazo. Os matemáticos Dave Bayer e Persis Diaconis estudaron sobre este asunto¹, centrándose nos métodos máis habituais de barallado e analizando o número de veces que é necesario, e suficiente, mesturar as cartas en cada caso para que estas queden ben baralladas. Un dos estilos analizados por Bayer e Diaconis foi a mestura americana, que consiste en separar a baralla en dous montóns aproximadamente iguais en número de cartas e ir intercalando naipes de ambos os montóns ata volver xuntar todas as cartas do mazo. Este estilo é un dos máis coñecidos e empregados, sobre todo nos casinos, pero cantas veces é preciso repetilo para que se produza unha verdadeira orde aleatoria nas cartas? Supoñamos que inicialmente temos as cartas da bara-

lla ordenadas de forma ascendente e realizamos unha mestura americana, neste proceso pasamos de ter unha secuencia de cartas crecente a ter dúas secuencias crecentes, facilmente identificables, dentro do mazo; situación que se atopa moi lonxe da aleatoriedade. Continuando co proceso, cunha nova mestura de cartas o número de secuencias crecentes volve a duplicarse; atopámonos un pouco máis preto do obxectivo buscado. En definitiva, cantas máis veces se realice este proceso máis próximos estaremos á situación de que as cartas se sitúen dentro do mazo de forma aleatoria. Segundo o traballo publicado polos citados autores, conclúese que o número óptimo de mesturas necesarias para garantir ese bo barallado, no caso da baralla inglesa, serían sete; menos, serían insuficientes e máis, serían innecesarias.

Chegados a este punto... COMECEMOS A PARTIDA!

¹ Bayer D, Diaconis P. Trailing the dovetails shuffle to its lair. The Annals of Applied Probability. 1992;2(2):294-313.

Dirección:

Paula Raña Míguez

Equipo editorial:

Esther Calvo Ocampo, M^a Jesús Casado Barrio, Balbina Casas Méndez, Tomás Cotos Yáñez, Rosa Crujeiras Casais, M^a Esther López Vizcaíno, Silvia M^a Lorenzo Freire, Javier Roca Pardiñas e Rebeca Villaverde López.

Revisión lingüística:

María Miragaya Pereira

Coa colaboración de:

Antonio Vaamonde Liste, Patricio Sánchez Fernández, M^a José Ginzo Villamayor, Miguel Ángel Rodríguez-Gigirey Pérez e Gael Naveira Barbeito.

Comunicación coa SGAPEIO:

Facultade de Matemáticas, Campus Vida, 15706-Santiago de Compostela, A Coruña.

www.sgapeio.es
secretaria@sgapeio.es
boletin@sgapeio.es
Twitter @sgapeio
LinkedIn SGAPEIO

COLABORA !!!

Informest é unha publicación cuatrimestral da SGAPEIO que podes atopar na web da sociedade, www.sgapeio.es, no apartado PUBLICACIÓN.

Se queres participar cun artigo de divulgación para a sección "Colaboracións" ou cunha curiosidade estatística para o "Sabías que?", contacta con nós no correo "boletin@sgapeio.es" onde te informaremos das normas de edición.

Anímate!!



SGAPEIO

Sociedade Galega para a Promoción da Estatística e da Investigación de Operacións