

In memoriam

Colaboración

Software

Actividades da SGAPEIO

Novas

Novas do IGE

Traballos de Estatística e IO

Coñecéndonos

Sabías que?

Editorial

Javier Tarrío Saavedra – Vicepresidente de Estatística

Xusto despois do cambio de hora, neste lento deitar de follas, ulindo xa as primeiras castañas, teño a oportunidade de presentar o novo número da nosa revista, editada desde comezos deste ano por Alejandro Saavedra Nieves e María Gómez Rúa (grazas María, Álex, por este fantástico traballo en parcería), e que inclúe as seccións de Colaboración, coa firma de Javier Roca, o Sabías que?, nesta ocasión da man de Covadonga Rodríguez Moldes, e o apartado de Software, escrito por Brais González, sen esquecer o Coñecéndonos, dedicado ao sobranceiro socio e compañeiro Jacobo de Uña, ademais dos habituais apartados de Novas do IGE e de Traballos de Estatística. Desde a SGAPEIO queremos expresar a todos o noso agradecemento polo seu labor desinteresado, así como tamén, de xeito especial, aos socios que toman nas mans a nosa revista para tranquilamente lela.

Queremos salientar que, tamén neste número, se publica unha sección moi sentida de homenaxe e lembranza ao noso compañeiro Alejandro Quintela del Río, recentemente falecido o pasado 3 de outubro. Investigador brillante, docente apaixonado e innovador, á parte do seu valioso traballo queda, nos amigos, nos compañeiros, nos seus alumnos e profesores, a lembranza e a pegada dos momentos compartidos cunha persoa única e excepcional, como así o testemuñan as súas palabras. A perda de Alejandro, unida á de Carmen Cadarso, deixa un pouso certamente amargo neste ano que termina.

Pero non quixera rematar este texto sen mencionar todo o que de excepcionalmente positivo depararon tamén estes últimos meses. De feito, todos os membros da SGAPEIO tomamos como alegrías propias os numerosos éxitos acadados polos seus socios, como é o caso de Ricardo Cao, nomeado *fellow* do prestixioso Institute of Mathematical Statistics e recente Medalla da Sociedad de Estadística e Investigación Operativa, seguindo a estela de Wenceslao González Manteiga, que igualmente a recibiu o pasado ano. Estes recoñecementos de excelencia profesional están tendo continuidade nos éxitos dos máis novos, gañadores no presente curso en dúas categorías diferentes da Fase Nacional da Incubadora de Sondaxes e Experimentos, sendo apenas a punta do iceberg do talento amosado por todos os alumnos e titores nas dúas fases do concurso.

E é precisamente este talento, o traballo e a transmisión do coñecemento un esquema que quizais podería describir o xeito de ser e facer da estatística galega. De entre todos estes tes, nos últimos tempos está cobrando especial relevancia a transmisión, como garante da necesaria continuidade e mellora na universidade e na sociedade en xeral, dirixida a un número cada vez maior de alumnos, tanto en titulacións relacionadas coa estatística como tamén nas demais carreiras universitarias. Ogallá este traballo continúe a prender a ilusión pola estatística e a investigación de operacións (pola ciencia en xeral) en novas xeracións de científicos; ogallá se dean cada vez máis oportunidades para cimentar carreiras e vocacións. Desde a SGAPEIO colaborarase sempre neste sentido, xa for a través dos PhDay ou na organización de cursos e premios para xoves investigadores, no marco dos nosos congresos. Precisamente, de aquí a un ano, aí polo Magosto, co cambio de hora, contamos con todos os socios para participar no próximo congreso da SGAPEIO (quede aquí constancia do noso agradecemento aos organizadores): Vémonos pronto na Coruña!

Dirección:

María Gómez Rúa
Alejandro Saavedra Nieves

Comunicación coa SGAPEIO:

www.sgapeio.es
secretaria@sgapeio.es
boletin@sgapeio.es
[@sgapeio](https://twitter.com/sgapeio)
[sgapeio](https://www.facebook.com/sgapeio)
[sgapeio](https://www.instagram.com/sgapeio)
[Sgapeio](https://www.linkedin.com/company/sgapeio)

Depósito Legal:

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

Lembranza ao noso compañeiro Alejandro Quintela del Río

por **Germán Aneiros Pérez e Graciela Estévez Pérez**
Departamento de Matemáticas da UDC



O pasado 3 de outubro recibimos a triste noticia do falecemento do noso compañeiro da área de Estatística e Investigación Operativa, o profesor Alejandro Quintela del Río. Tras máis de 18 meses de loita, a enfermidade gañoulle a batalla, a pesar do seu optimismo e vontade por recuperarse.

Alejandro Quintela licenciouse en Ciencias Matemáticas pola Universidade de Santiago de Compostela no ano 1986. Nada máis finalizar os seus estudos de licenciatura, comezou a súa carreira universitaria como profesor da área de Estatística e IO no entón Campus da Coruña da USC, defendendo a súa tese doutoral en 1992, baixo a dirección de Juan Vilar. No ano 1993 obtivo a titularidade como profesor do departamento de Matemáticas, na xa Universidade da Coruña, e foi na súa toma de posesión como catedrático, no ano 2012, cando Alejandro viu recompensada a súa gran dedicación á docencia e á investigación.

Foi socio fundador da SGAPEIO no ano 1993, director do

Departamento de Matemáticas da UDC e, actualmente, era o coordinador do Programa Oficial de Doutoramento en Saúde, Discapacidade, Dependencia e Benestar, adscrito á Facultade de Fisioterapia. Dirixiu as nosas teses de doutoramento, colaborou en investigación cun bo número de colegas estatísticos, tanto da UDC como doutras universidades españolas e estranxeiras, como o profesor Philippe Vieu da Université Paul Sabatier-Toulouse III, co cal mantivo sempre unha magnífica amizade.

Nos derradeiros anos centrou o seu labor docente e investigador no ámbito das Ciencias da Saúde, concretamente na Facultade de Fisioterapia, na que participou de forma moi activa en proxectos e actividades investigadoras do Grupo de Investigación Intervención Psicolóxica e Rehabilitación Funcional, do que formaba parte. Era unha persoa moi querida e prezada neste centro.

Máis aló de todo o seu traballo e méritos acadados, lembraremos a súa honradez e humanidade, e moitas outras boas calidades, como recollen os testemuños dalgúns compañeiros/as:

“Lémbroo como un alumno excelente e cun gran sentido do humor... moi traballador e moi consecuente coas súas ideas. A pesar de non ter unha vida doada, entre outras cousas polos seus frecuentes problemas de saúde, non perdía o sentido do humor.”

“Tiña unha mente brillante e sempre loitou polas súas conviccións.”

“Era un gran compañeiro e amigo, divertido e enxeñoso. Tiña os seus momentos complicados coma todo o mundo, pero sabías que sempre estaba aí para animarte e botarche unha man. Resultaba moi ameno traballar con el, mesmo en longas xornadas. Era moi resolutivo e eficaz. A verdade, unha gran sorte ter podido colaborar con Alejandro.”

“Coñecín a Alejandro no primeiro ano de carreira de Matemáticas onde xa despuntaba como estudante brillante e excelente compañeiro, de carácter alegre, sempre cunha broma que alixeiraba o traballo. Posteriormente, volví a compartir con el o traballo na Universidade no que seguiu destacando na innovación educativa e na tenacidade no

traballo. Quedará sempre na nosa memoria.”

“Lembro o seu humor, moitas veces acedo, sempre enxeñoso, que adoitaba utilizar para quitarlle ferro aos contratempos, xa fosen do ámbito persoal ou do académico. Estando con el sempre había momentos para as risas. Teño moitos motivos para estarlle agradecido.”

“Desde o primeiro día que coñecín a Alejandro, sendo el director do departamento de Matemáticas, sempre me axudou en todo. Tiven a sorte de colaborar con el en temas de investigación e recordo con agarimo aquelas longas tardes nas que falabamos non só de estatística, senón tamén de cinema e de política, temas que tanto lle gustaban a Alejandro e que enxeñosamente e con moito humor utilizaba nos seus libros e apuntamentos para espertar o interese do alumnado. Alejandro, sempre che

estarei agradecido. Botareite de menos.”

“I enjoyed my collaborations with Alejandro. Beyond his great professional qualities, I remember his kindness and his constant humanity that made him so endearing.”

“Alejandro restera à la fois un ami et un compagnon de travail; un de ceux que l'on n'oublie jamais. Les heures de calculs interminables, les soirées entre amis à décortiquer «percebes» et «zamburiñas», les balades à travers la Galice que j'ai découvert grâce à lui pendant plus de 30 années et ses séjours à Toulouse resteront gravés dans ma mémoire. Son humour à toute épreuve manque déjà.”

Benquerido Alejandro, sempre permanecerá entre nós a túa lembranza. Descansa en paz.



Alejandro Quintela del Río.

Aplicación da regresión segmentada á análise do impacto da pandemia de SARS-CoV-2 nas infeccións e uso de antibióticos na poboación pediátrica da área de Vigo

por Ana Clavería Fontán e Javier Roca Pardiñas

Ana Clavería é doutora en Medicina Preventiva e MBA, e pertence ao nodo de Galicia da Rede de Investigación en Cronicidade, Atención Primaria e Promoción da Saúde. Participa na Plataforma de Medicina Preditiva como Co-IP da Comisión de Historia Clínica e é a investigadora principal do [Grupo I-Saúde](#) do Instituto de Investigación Sanitaria Galicia-Sur. Tamén é responsable da investigación en Atención Primaria da EOXI de Vigo, e participa activamente na European General Practice Research Network, como representante española na Rede.

Javier Roca Pardiñas é doutor en Estatística e Investigación Operativa pola Universidade de Santiago de Compostela. Actualmente é catedrático de universidade no Departamento de Estatística e Investigación Operativa da Universidade de Vigo e membro do grupo SiDOR da mesma universidade. Na actualidade é o director adxunto de transferencia do Centro de Investigación e Tecnoloxía Matemáticas de Galicia (CITMAga).

Os deseños tradicionais de estudos epidemiolóxicos, como os estudos de cohortes e casos-control, poden proporcionar evidencia importante sobre a etioloxía da enfermidade, pero son menos útiles como estudos de intervención, debido a limitacións como a confusión por diferenzas de grupo e, en particular, ao nesgo dos usuarios sans. O *gold standard* para a xeración de evidencia relacionada coas intervencións sanitarias, diagnósticas, clínicas, organizativas e educativas é o ensaio clínico aleatorizado (ECA). A teoría detrás desta metodoloxía é que os posibles prexuízos relacionados coa heteroxeneidade do paciente e as covariables de confusión distribúense uniformemente entre os grupos estudados e, polo tanto, non inflúen de forma diferente no tratamento. En termos estatísticos, as probas en deseños de ECA para a diferenza de resultados entre dous grupos baséanse no cálculo da diferenza entre os expostos ao tratamento e os que non, ignorando por completo a variabilidade sub-xacente. Non obstante, os ECA teñen un alcance limitado nas intervencións comunitarias, xa que non é factible, e ás veces non é ético, aleatorizar a intervención. Por deseño, os ECA explicativos non teñen en conta nin poden ter en conta a variedade de dimensións da demografía dos pacientes, as variacións na saúde e a complexidade xeral

da atención sanitaria.

As series temporais interrompidas (*Interrupted Time Series*, ITS) ofrecen unha metodoloxía rigorosa para determinar a eficacia de intervencións sanitarias complexas en contextos reais. As análises de ITS son un dos deseños avaliativos máis fortes cando a aleatorización non é posible. Son especialmente útiles cando se producen "experimentos naturais", por exemplo cando se produce un cambio de política sanitaria. As súas principais vantaxes fronte aos enfoques alternativos é que poden facer pleno uso da natureza lonxitudinal dos datos, tendo en conta as tendencias previas á intervención; e que, cando se realizan en contornas do mundo real, mellora a súa validez externa. Outra característica importante para garantir que a investigación se traduza na práctica é que a presentación gráfica e numérica dos resultados pode ser facilmente entendida por aqueles con pouco coñecemento experto en métodos estatísticos e epidemiolóxicos. Polo tanto, cando os ECA non son viables ou aplicables, os ITS son cada vez máis recoñecidos como aplicables e considéranse o suficientemente rigorosos como para incluírse nas metaanálises Cochrane [4, 3]. En esencia, os ITS son particularmente útiles cando un ensaio aleatorio é inviable ou non é ético.

O deseño do estudo da serie temporal interrompida (ITS) utilízase cada vez máis para avaliar intervencións de saúde pública; é especialmente axeitado para intervencións baseadas na poboación durante un período de tempo claramente definido e con resultados de saúde no nivel de poboación obxectivo. O ITS utilízase para a avaliación dunha ampla gama de intervencións de saúde pública, incluíndo novas vacinas, o impacto dos cascos de bicicleta obrigatorios, cambios nos envases de paracetamol, zonas de velocidade de tráfico e precaucións contra as infeccións nosocomiais, así como na avaliación do impacto sobre a saúde de eventos non planificados como a crise financeira global ou unha pandemia global como o Covid-19.

No grupo I-Saúde, do Instituto de Investigación Sanitaria Galicia Sur, de composición multidisciplinar (médicas, pediatras, epidemióloga, matemática, enfermeira) aplicamos esta técnica estatística para analizar as enfermidades infecciosas pediátricas e o consumo de antibióticos nunha rexión atlántica europea durante a pandemia de SARS-CoV-2 [2].

Escollemos este tema pola crecente preocupación pola resistencia bacteriana, que fixo aínda máis urgente a prescrición racional de antibióticos. As medidas non farmacolóxicas establecidas para reducir o impacto da pandemia de SARS-CoV-2 modificaron a epidemioloxía das infeccións pediátricas e, en consecuencia, o uso de antibióticos. Coa análise de ITS cuantificouse e analízase o efecto da pandemia de SARS-CoV-2 na incidencia de enfermidades infecciosas e no uso de antibióticos entre 2018 e 2020 na Área Sanitaria de Vigo (Galicia, España). Trátase dun estudo de datos reais con rexistros administrativos dos servizos de atención primaria prestados a toda a poboación pediátrica da área de Vigo. Os rexistros relacionáronse con episodios clasificados como infecciosos pola Clasificación Internacional de Atención Primaria (ICPC-2) e con medicación oral no subgrupo terapéutico J01, correspondentes a antibióticos de uso sistémico, segundo a clasificación ATC (Sistema de Clasificación Anatómica, Terapéutica, Química da OMS). Clasifícanse os rexistros segundo episodios de incidencia, idade, dose por habitante e ano.

O método estatístico empregado permite o estudo dunha serie temporal de interese ao longo do tempo. En par-

ticular, denotando por X_t o valor da devandita serie no instante t , consideraremos o modelo de regresión $X_t = m(t) + e_t$ onde a función $m(\cdot)$ representa a tendencia da serie e e_t é o erro temporal correspondente. Neste contexto, utilízase Metodoloxía ARIMA ou Box-Jenkins como técnica estatística, que permite a obtención de modelos matemáticos que establecen a estrutura temporal (tendencia, estacionalidade,...) da serie. Este tipo de modelos son moi útiles para a previsión a curto prazo. Por outra banda, considérase que a tendencia $m(t)$ pode ser axustada de forma lineal en cada K intervalos de tempo consecutivos:

$$I_1 = (t_0, t_1], I_2 = (t_1, t_2], \dots, I_K = (t_{K-1}, t_K]$$

Neste tipo de modelos, é necesario estimar a partir dos datos mostrais, tanto o número de intervalos K como a posición dos puntos de corte (t_0, t_1, \dots, t_K) , que determinan as seguintes relacións:

$$m(t) = \begin{cases} a_1 + b_1 \cdot X_t & \text{si } t \in I_1 = (t_0, t_1] \\ a_2 + b_2 \cdot X_t & \text{si } t \in I_2 = (t_1, t_2] \\ \vdots & \\ a_K + b_K \cdot X_t & \text{si } t \in I_K = (t_{K-1}, t_K] \end{cases}$$

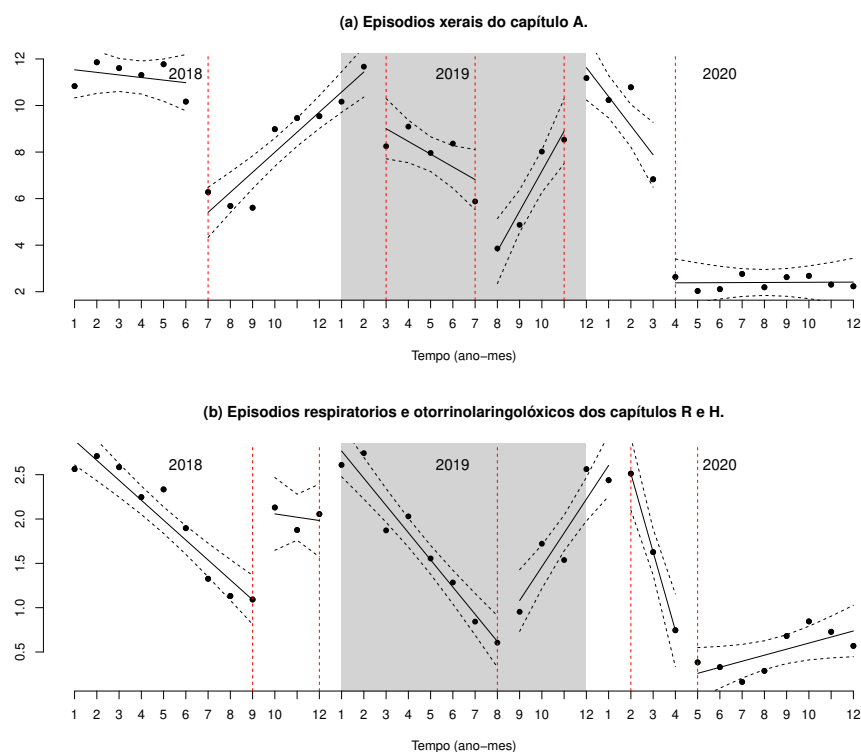


Figura 1: Regresión segmentada do número de episodios engadidos. (a) Episodios xerai, incluídos no capítulo A da CIAP-2. (b) Episodios respiratorios e otorrinolaringolóxicos, correspondentes aos diagnósticos dos capítulos R e H da CIAP-2. A CIAP-2 é a Clasificación Internacional de Atención Primaria, unha clasificación da terminoloxía médica de ámbito internacional.

A aplicación deste tipo de modelos de regresión segmentada revelou un efecto significativo da pandemia de SARS-CoV-2 sobre os episodios infecciosos rexistrados nas historias clínicas de atención primaria nunha poboación pediátrica. Ao mesmo tempo, produciuse unha importante diminución da prescrición de antibióticos, especialmente penicilinas, cefalosporinas e macrólidos.

A Figura 1 presenta a evolución mensual, sumando os diagnósticos respiratorios e otorrinolaringolóxicos, os diagnósticos do capítulo A, que inclúen principalmente enfermidades víricas exantemáticas (sarampelo, varicela, rubéola, febre) e SARS-COV-2. Nesta figura obsérvanse 7 puntos de cambio, cun significativo salto en setembro de 2020 ao comezo do curso escolar cando se dispoñían das probas diagnósticas para SARS-CoV-2. Os episodios respiratorios e otorrinolaringolóxicos presentaron seis puntos de cambio, e a regresión segmentada correspondente ao período de confinamento nos primeiros meses de 2020 mostra un marcada tendencia descendente, seguida por outra máis horizontal.

A Figura 2 mostra a evolución das penicilinas e dos ma-

crólidos. No consumo de antibióticos (penicilinas, cefalosporinas e macrólidos) identificáronse cinco puntos significativos de cambio. Os cambios de constantes e pendentes para as regresións segmentadas identificadas seguiron unha evolución similar á observada nas enfermidades respiratorias e otorrinolaringolóxicas, cunha tendencia á baixa moi acusada a principios de 2020, seguida dun pequeno cambio nos meses seguintes.

Como punto forte desta metodoloxía, cabe destacar a especificación do intervalo de tempo (meses), o resultado (proporción ou DHD), o modelo de forma (cambio de nivel e cambio de pendente), a aplicación da autocorrelación e as medidas de efecto (cambio de nivel e cambio de pendente, incluíndo p-valores e intervalos de confianza) [1, 5]. Xa temos iniciada a aplicación desta técnica á análise da mortalidade por SARS-Cov2, e a colaboración con grupos de investigación da Red de Cronicidad, Atención Primaria y Promoción de la Salud (RICAPPS) do Instituto de Salud Carlos III para avaliar diversas intervencións en saúde pública.

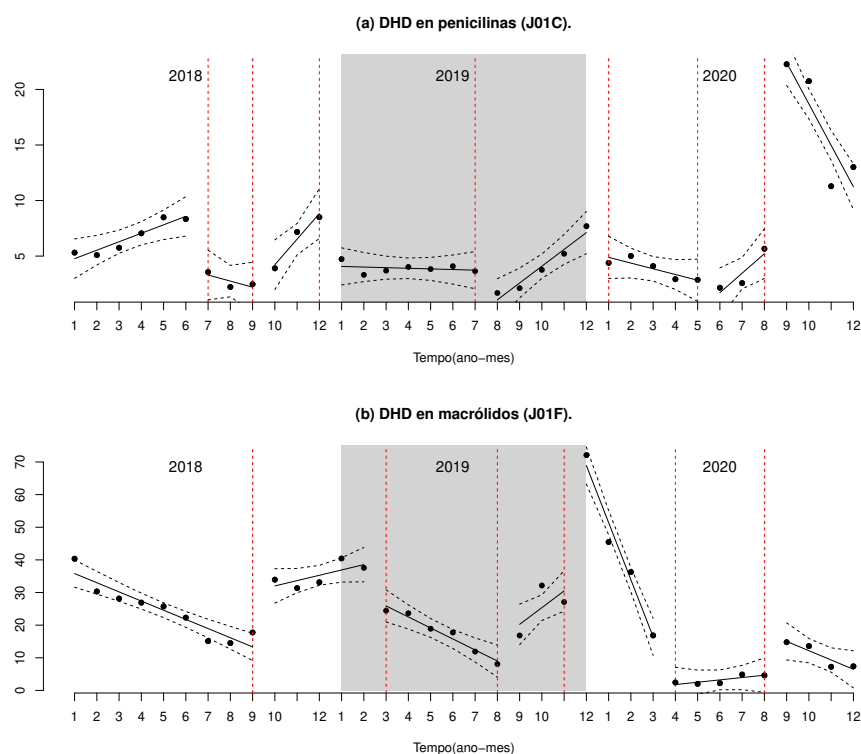


Figura 2: Regresión segmentada do número de doses habitante por día (DHD). (a) DHD en penicilinas (J01C). (b) DHD en macrólidos (J01F).

Referencias

- [1] Bernal JL, Soumerai S, Gasparrini A. A methodological framework for model selection in interrupted time series studies. *Journal of Clinical Epidemiology* 2018; 103: 82-91
- [2] Clavería A, Delgado-Martín MV, Goicoechea-Castaño A, Iglesias-Moreno JM, García-Cendón C, Martín-Miguel MV, . . . , Roca-Pardiñas J.. Interrupted time series analysis of pediatric infectious diseases and the consumption of antibiotics in an Atlantic European region during the SARS-CoV-2 pandemic. *Antibiotics* 2022; 11(2): 264.
- [3] Hategeka C, Ruton H, Karamouzian M, Lynd LD, Law MR. Use of interrupted time series methods in the evaluation of health system quality improvement interventions: A methodological systematic review. *BMJ Glob. Health* 2020; 5(10): e003567.
- [4] Polus S, Pieper D, Burns J, Fretheim A, Ramsay C, Higgins JP, Mathes T, Pfadenhauer LM, Rehfues EA. Heterogeneity in application, design, and analysis characteristics was found for controlled before-after and interrupted time series studies included in Cochrane reviews. *Journal of Clinical Epidemiology* 2017; 91: 56-69.
- [5] Turner SL, Karahalios A, Forbes AB, Taljaard M, Grimshaw JM, Cheng AC, . . . , McKenzie JE. Design characteristics and statistical methods used in interrupted time series studies evaluating public health interventions: a review. *Journal of Clinical Epidemiology* 2020; 12: 1-11.

RAPOSa: unha ferramenta para resolver problemas de optimización polinómica

por Brais González Rodríguez

RAPOSa (Reformulation Algorithm for Polynomial Optimization - Santiago) é unha ferramenta para resolver problemas de optimización polinómica. Os problemas de programación polinómica son de gran relevancia na práctica, pois serven para modelar diversas situacións como problemas de control óptimo, problemas de xestión de redes de enerxía, problemas de localización ou situacións derivadas das finanzas. Por outra parte, os problemas de optimización lineal con variables binarias ou os problemas cadráticos, ambos de ampla relevancia na práctica, pódense expresar como casos concretos dos problemas de programación polinómica.

O algoritmo detrás de RAPOSa baséase no que se coñece como Reformulation-Linearization Technique (de agora en adiante RLT), introducida en [2]. A continuación explicaremos brevemente en que consiste dita técnica. Para isto, necesitamos definir primeiro o que entendemos por problema de optimización polinómica, tal e como se mostra a continuación.

$$\begin{aligned} \min \quad & \phi_0(\mathbf{x}) \\ \text{suxeito a} \quad & \phi_r(\mathbf{x}) \geq \beta_r, \quad r = 1, \dots, R_1 \\ & \phi_r(\mathbf{x}) = \beta_r, \quad r = R_1 + 1, \dots, R \\ & \mathbf{x} \in \Omega \subset \mathbb{R}^n, \end{aligned}$$

($PP(\Omega)$)

onde $\Omega = \{\mathbf{x} \in \mathbb{R}^n : 0 \leq l_j \leq x_j \leq u_j < \infty, \forall j \in N\} \subset \mathbb{R}^n$ é un hiperrectángulo e $\phi_r(\mathbf{x})$ son polinomios

de grao δ_r , para todo $r = 0, \dots, R$. Por outra parte, definimos o grao do problema como $\delta = \max_{r=0, \dots, R} \delta_r$ e denotamos por $N = \{1, \dots, n\}$ os índices de cada variable.

A técnica RLT consiste en construír relaxacións lineais do problema ($PP(\Omega)$). Para iso, defínese unha variable RLT para cada monomio de grao maior que 1. Para cada multiconxunto¹ de índices J , definimos a variable RLT $X_J = \prod_{j \in J} x_j$. Desta forma, dado un polinomio $\phi(\mathbf{x})$ definimos a súa linealización $[\phi(\mathbf{x})]_L$ como o resultado de substituír en $\phi(\mathbf{x})$ todos os monomios de grao maior que 1 pola súa correspondente variable RLT. Por outra parte, necesitamos definir o que se coñece como bound-factor constraints, que son da forma $F_\delta(J_1, J_2) = \prod_{j \in J_1} (x_j - l_j) \prod_{j \in J_2} (u_j - x_j) \geq 0$, onde $J_1 \cup J_2 \subset N^\delta$ e $|J_1 \cup J_2| = \delta$.

Con estes ingredientes, construímos a seguinte relaxación lineal do problema $PP(\Omega)$ da seguinte forma:

$$\begin{aligned} \min \quad & [\phi_0(\mathbf{x})]_L \\ \text{suxeito a} \quad & [\phi_r(\mathbf{x})]_L \geq \beta_r, \quad \text{para cada } r = 1, \dots, R_1 \\ & [\phi_r(\mathbf{x})]_L = \beta_r, \quad \text{para cada } r = R_1 + 1, \dots, R \\ & [\prod_{j \in J_1} (x_j - l_j) \prod_{j \in J_2} (u_j - x_j)]_L \geq 0, \\ & \text{para cada } \begin{matrix} J_1 \cup J_2 \subset N^\delta, \\ |J_1 \cup J_2| = \delta \end{matrix} \\ & \mathbf{x} \in \Omega \subset \mathbb{R}^n. \end{aligned}$$

($LP(\Omega)$)

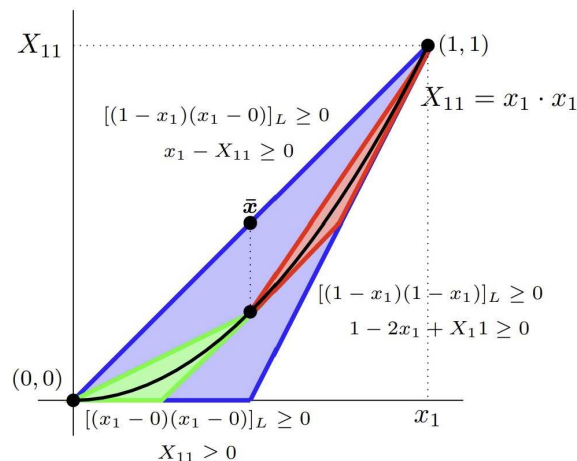


Figura 3: Representación gráfica da ramificación usando a técnica RLT

¹Por multiconxunto enténdese a extensión do concepto de conxunto incluíndo elementos repetidos.

²Denotamos por N^δ o multiconxunto $\{1, \dots, 1, \dots, n, \dots, n\}$.

O algoritmo detrás de RAPOSa consiste en construír o problema $LP(\Omega)$ e, a partir da súa solución, utilizar o método de ramificación e acoutación, xerando subproblemas lineais coa rexión factible cada vez máis pequena. O criterio para elixir a variable de ramificación baséase nas violacións das ecuacións que definen as variables RLT ($X_J = \prod_{j \in J} x_j$), xa que no momento no que se cumpran estas ecuacións, a solución da relaxación correspondente será factible no problema polinómico $PP(\Omega)$. Na Figura 3 móstrase o que fai este algoritmo na súa primeira ramificación para un problema sinxelo de minimizar unha certa función obxectivo sobre a curva x_1^2 , con $0 \leq x_1 \leq 1$.

Finalmente, cabe destacar que a ferramenta RAPOSa está dispoñible para utilizar de forma gratuíta. Po-

de descargarse dende a páxina web <https://raposa.usc.es/>. Na propia páxina web atópase información sobre como resolver problemas de optimización polinómica utilizando RAPOSa. A modo de resumo, mencionar que é compatible coas linguaxes AMPL, Pyomo e Julia, as tres moi utilizadas para resolver problemas de optimización matemática. O usuario tamén pode executar RAPOSa directamente no servidor NEOS, sen necesidade de descargar o executable, na páxina web <https://neos-server.org/neos/solvers/go:RAPOSa/AMPL.html>. Por último, no artigo [1] pode atoparse unha descrición máis detallada do funcionamento de RAPOSa, así coma unha análise computacional detallada do seu rendemento, unido a diferentes melloras do algoritmo.

Referencias

- [1] Gonzalez-Rodríguez, B., J. Ossorio-Castillo, J. González-Díaz, Á. M. González-Rueda, D. R. Penas, and D. Rodríguez-Martínez. Computational advances in polynomial optimization: RAPOSa, a freely available global solver. *Journal of Global Optimization* 2022. In press.
- [2] Sherali, H. D. and C. H. Tuncbilek): "A global optimization algorithm for polynomial programming problems using a Reformulation-Linearization Technique", *Journal of Global Optimization* 1992; 2, 101-112.

XII Edición do Concurso Incubadora de Sondaxes e Experimentos e Fase Nacional

por **María Martín Vila, M^a Dolores Rico Ramos, Covadonga Rodríguez-Moldes Rey e Olga Simón Cruz**

Na duodécima edición do Concurso da Incubadora de Sondaxes e Experimentos presentáronse un total de 41 traballos distribuídos por categorías do seguinte xeito: de 1^º e 2^º da ESO 16 traballos, de 3^º e 4^º da ESO 22 traballos e da categoría de Bacharelato e Ciclos Formativos de Grao Medio 3 traballos.

En canto á temática, foi moi variada e non se centrou tanto na COVID-19 e as súas consecuencias como nos dous últimos anos e recuperamos así a diversidade ademais da ansiada normalidade. Nos traballos presentados na categoría de 1^º e 2^º da ESO, tratáronse temas como o galego, a normalidade, as preferencias sobre rural ou urbano, a cultura, as marcas e a súa aceptación, o impacto da pandemia nos gastos ou o camiño de Santiago.

Na categoría de 3^º e 4^º da ESO abordouse o tema da saúde dende distintos puntos de vista, desde a saúde mental ata a súa relación cos hábitos de vida, a nutrición ou o exercicio. Tamén se trataron temas como a publicidade en internet, a relación entre as actividades extraescolares e as notas, os xogos de azar, o teito de cristal ou a actividade dos animais no Courel.

Por último, na categoría de Bacharelato e Ciclos, presentáronse estudos sobre a eficacia das vacinas da COVID-19, a evolución no rol das mulleres ou o estudo da densidade de poboación dun barrio.

Cabe destacar o uso de diferentes formas de recollida de datos: en papel, empregando formularios de google docs ou cuestionarios en follas de cálculo. Para o seu posterior tratamento, empregouse principalmente unha folla de cálculo. Como en anteriores edicións, dende a SGA-

PEIO, queremos agradecer a todos os participantes o labor desenvolvido: ao alumnado, pola dedicación, esforzo e talento na consecución do traballo e, aos seus titores ou titoras, por espertar a curiosidade pola Estatística e a Investigación de Operacións como ferramenta para resolver problemas que xorden en multitude de campos. O xurado manifestou a alta calidade de todos os traballos presentados a este concurso.

Os traballos gañadores son:

- Categoría 1^º e 2^º da ESO: “XACOMAT 2022, o inicio dun gran camiño” dos alumnos Julia Bello Gómez, Ruyi Ye, Sofía Seoane Recacho, Fabio Antelo Reza e Óscar Rozas Midón, titorizados polo profesor Manuel Cabaleiro Calveiro do Colegio San José de Cluny (Santiago de Compostela).
- Categoría 3^º e 4^º da ESO: “Estudo da relación entre os casos de covid-19 e a contaminación do aire nas cidades de Vigo e Valladolid” dos alumnos Iván Iglesias Campos, Antón Justo Vicente, Nora Ramalla Martínez e Sara Vicente Lorenzo, titorizado polos profesores Adela Rosa Rodríguez Rodríguez e Daniel Fernández Vence do IES República Oriental do Uruguai (Vigo).
- Categoría Bacharelato e Ciclos Formativos de Grao Medio: “Onde caben dous caben cinco” dos alumnos Pablo Araújo Rodríguez, Florencia Isabel Balmelli Rivero, Eduardo A. Tella Vidal e Isabel Rodríguez Políno, titorizados pola profesora María Ángel Martínez Rodríguez do IES Agra do Orzán (A Coruña).



Participantes da fase local en Santiago.

Os traballos finalistas resultaron:

- Categoría 1º e 2º da ESO: “É posible aprobar un exame facéndoo ao chou?” dos alumnos Anxo Alcalde Iglesias, Érika Conde Pereiro, Leo Landeira Blanco e Alejandro Ríos Ferreiro, titorizados polo profesor Fabio José González Eiras do CPI de Bembibre (A Coruña).
- Categoría 3º e 4º da ESO: “Un ano no Courel” dos alumnos Pablo Álvarez Marcos, Zaira Arza González e Sabela Rodríguez Rodríguez, titorizado pola profesora Lourdes González Sotelo do CPI Poeta Uxío Novoneyra (Lugo) e “¿Cómo afectan determinados factores al grosor del pelo?” das alumnas Ainhoa Caride Estévez, Ainhoa Caballero Pérez, Laura Gómez Domínguez e Kiara Vázquez Lescano, titorizados pola profesora Mariña López Yunta do IES Ribeira do Louro (O Porriño).
- Categoría Bacharelato e Ciclos Formativos de Grao Medio: “Situación das mulleres en diferentes momentos temporais” dos alumnos Paula Álvarez López, Carla Díaz Rodríguez, Alejandro Fernández Brea, Cati Fernández Pardo e Iván Vila Carrete, titorizados pola profesora Fe Pérez Pérez do IES Gregorio Fernández (Lugo).

O xurado decidiu outorgar mencións aos seguintes traballos:

- Categoría 1º e 2º da ESO: “Como caen os zapatos?” dos alumnos Diego Blanco Costa, Julia Calvo Suárez, Tania Cancela Cerviño e Samuel Gerpe Francisco, “As marcas inflúen nos gustos?” dos alumnos Sergio Blanco Vázquez, Iván García Caamaño, Sabrina Lois Goimil, Pablo Migueles Grille e Adrián Rey López, ambos titorizados polo profesor Fabio José González Eiras do CPI de Bembibre (A Coruña).
- Categoría 3º e 4º da ESO: “O teito de cristal” dos alumnos Telma Campos Fernández, Simón Cano Reboreda, Pedro Rodríguez Fernández e Mateo Álvarez Pons, titorizados polas profesoras María José Lorenzo Fernández e Adela Rosa Rodríguez Rodríguez do IES República Oriental do Uruguay (Vigo) e “¿Las personas con gafas sacan mejores notas?” dos alumnos Sergio González Martínez, Leo Salgueiro Prieto, Diego Romero Domínguez e Izan José Rodríguez Martínez, titorizados pola profesora Mariña López Yunta do IES Ribeira do Louro (O Porriño).

O xurado deste *XII Concurso Incubadora de Sondas e Experimentos* estivo composto por Covadonga Rodríguez-Moldes Rey (SGAPEIO), que actuaba en calidade de Presidenta; Javier Tarrío Saavedra (SGAPEIO),

como secretario; e como vogais Indalecio Cabana Leira (Consellería de Cultura, Educación, Formación Profesional e Universidades), María Estefanía Villar Cheda (Instituto Galego de Estatística), Leticia Lorenzo Picado (SGAPEIO), como representante de CITMAGa, e Olga Simón Cruz (SGAPEIO).

A decisión final dos gañadores desta edición e entrega de premios celebrouse o 14 de xuño na Aula Magna da Facultade de Matemáticas. No acto de clausura e entrega de premios, estiveron presentes a decana da Facultade de Matemáticas, Dona Elena Vázquez Cendón, o subdirector de produción Estatística do Instituto Galego de Estatística, Don Manuel Mateo Díaz, o director da SGAPEIO, Don César Andrés Sánchez Sello, a presidenta do xurado e vogal da SGAPEIO, Dona Covadonga Rodríguez-Moldes Rey e o Delegado territorial da Consellería de Cultura, Educación, Formación Profesional e Universidades, Don Indalecio Cabana Leira.

O concurso nesta edición conta coa colaboración da Facultade de Matemáticas da USC, a Consellería de Cultura, Educación, Formación Profesional e Universidades o Instituto Galego de Estatística, o Centro de Investigación e Tecnoloxía Matemática de Galicia (CITMAGA) e Gadis.

Fase nacional

Os traballos gañadores foron elixidos para participar na [FASE NACIONAL 2021/2022](#), celebrada en Toledo do día 23 ao 25 de xullo, en representación da Comunidade Galega. O traballo presentado “XACOMAT 2022, o inicio dun gran camiño?” resultou gañador na categoría 1º e 2º da ESO e o traballo “Onde caben dous caben cinco” resultou gañador *ex aequo*, xunto con outros tres, na categoría de Bacharelato e Ciclos Formativos de Grao Medio.



Equipos que se presentaron á fase nacional en Toledo.



INSTITUTO GALEGO DE
ESTADÍSTICA

Novidades na información estatística publicada

Indicadores migratorios

Os “Indicadores migratorios” do IGE toman como espazo territorial de referencia a comunidade autónoma de Galicia; consideran movemento migratorio interior aquel que teña orixe e destino un concello galego, e movemento migratorio exterior todo movemento que teña como orixe ou destino o resto de España ou o estranxeiro. Estes movementos clasifícanse segundo xénero, grupos de idade e nacionalidade, e a partir deles calcúlanse saldos migratorios, taxas de emigración/inmigración ou idades medias á migración. Inclúense tamén outras variables relacionadas co fenómeno migratorio, que aínda que non miden directamente movementos migratorios, proporcionan información relevante sobre o seu desenvolvemento, como son as obtidas dos datos sobre residentes estranxeiros en Galicia ou de persoas de nacionalidade española con vinculación cun concello galego residentes no estranxeiro.

[+ info](#)

Posición cíclica da economía galega

A actual normativa europea sobre marcos orzamentarios establece a necesidade de introducir a planificación pluri-anual na política económica de todas as administracións públicas baseada nos principios de estabilidade orzamentaria e sustentabilidade financeira. O seu cumprimento en España desenvólvese coa Lei Orgánica 2/2012, do 27 de abril, de Estabilidade Orzamentaria e Sustentabilidade Financeira, en Galicia pola Lei 2/2011, do 16 de xuño de 2011, de Disciplina Orzamentaria e Sustentabilidade Financeira.

O principio de estabilidade orzamentaria conséguese cando unha administración pública non incorre en déficit estrutural. Ademais, o gasto non poderá aumentar por enriba dun valor de referencia, que no caso de España estableceuse para todas as administracións públicas na taxa de crecemento potencial do Produto interior bruto do conxunto de España.

Neste contexto, xorde a necesidade de estimar o crecemento potencial e o *output gap* da economía galega cun dobre obxectivo. Por unha parte, comparar o crecemento potencial da economía galega coa española, e por outra estimar a descomposición do déficit en cíclico e estrutural.

[+ info](#)

Carga de poboación estacional dos concellos

O Instituto Galego de Estatística (IGE) publica a actividade estatística Carga de poboación estacional dos concellos de Galicia, cuxo obxectivo é cuantificar a poboación que soportan os concellos en cada trimestre do ano.

A estimación da carga de poboación concíbese como unha estatística de síntese que aproveita as fontes estatísticas existentes sobre a poboación residente e as persoas que se desprazan por motivos de turismo, estancias en segundas vivendas, estudos, traballo e problemas de saúde.

Esta información sobre a carga de poboación pode ser útil para a toma de decisións en cuestións como a dotación de servizos sanitarios, planificación do transporte público, etc.

O ano obxecto de análise, ano 2020, está marcado significativamente pola crise da COVID-19, coincidindo o segundo trimestre co período de confinamento obrigatorio declarado no estado de alarma. Aínda que o segundo trimestre é o máis afectado, as consecuencias en canto á mobilidade das persoas estendéronse ata o final do ano. Para adecuar o estudo a esta situación de pandemia fixéronse algunhas modificacións no cálculo das compoñentes do turismo, traballo e estudos que poden consultarse na metodoloxía.

[+ info](#)

Convenios e acordos

- Acordo de colaboración entre a Consellería do Medio Rural, a Axencia Galega da Industria Forestal e o Instituto Galego de Estatística para o intercambio de datos e información (xullo 2022).
- Convenio de colaboración sobre intercambio de información entre a Universidade de Vigo e o Instituto Galego de Estatística para desenvolver actuacións do Plan Galego de Estatística 2022-2026 relacionadas co mundo académico (xullo 2022).
- Convenio de colaboración sobre intercambio de información entre a Universidade da Coruña e o Instituto Galego de Estatística para desenvolver actuacións do Plan Galego de Estatística 2022-2026 relacionadas co mundo académico (setembro 2022).

Traballos defendidos no SUG

Teses

New approaches to nonparametric circular regression models

María Alonso Pena.

Directora: Rosa María Crujeiras Casais.

Data da defensa: 8 de setembro de 2022. USC

A dirección na que un animal escapa dunha ameaza é un exemplo de variable aleatoria con soporte na circunferencia unidade. Outros exemplos inclúen direccións de vento, das ondas do mar ou ángulos de fracturas en ósos. Estas observacións periódicas son coñecidas como datos circulares, e constitúen un caso particular de estruturas de datos máis complexas denominadas datos direccionais.

Os métodos estatísticos clásicos non resultan axeitados para o correcto modelado destas observacións, e a súa análise estatística debe ter en conta a natureza periódica das variables, que non debería depender da dirección orixe seleccionada ou do sentido de rotación elixido.

O número de traballos na literatura que tratan a análise de datos circulares ten aumentado de xeito considerable nos últimos anos. Un problema interesante é o estudo da dependencia entre unha variable aleatoria circular e outras variables de diferente tipo, que pode ser tratado dende a perspectiva da regresión. O enfoque máis común para modelar funcións de regresión, normalmente consideradas como a media condicional, é a suposición de formas paramétricas e a posterior estimación dos parámetros do modelo. Porén, poden obterse vías de estimación máis flexibles empregando métodos non paramétricos, como os tipo núcleo.

O obxectivo desta tese é a proposta e estudo de estimadores tipo núcleo para a función de regresión cando, ao menos, unha das variables de interese presenta natureza circular. A función de regresión considérase, primeiramente, como a media condicional da variable resposta, dado o valor da covariable. Tamén se estuda un escenario máis xeral, onde as curvas de regresión son conceptualizadas como características xerais da distribución condicional, e a estimación realízase mediante máxima verosimilitude local. Finalmente, lévase a cabo a estimación das modas locais condicionais, establecendo o tratamento da regresión multimodal circular. Para concluír, estúdase o funcionamento dos modelos considerados e estes son aplicados a datos reais.

Traballos Fin de Máster

Desenvolvemento dun Priorizador para Campañas Comerciais Empregando Modelos Preditivos

José Damián García Noya

Directores: Rubén Fernández Casal e Guillermo López Taboada.

Titor: Juan Manuel Mazaira Gómez (ABANCA)

UDC. Modalidade prácticas

Unha das liñas de traballo actuais na área de Intelixencia de Clientes da entidade financeira ABANCA Corporación Bancaria S.A é o desenvolvemento dun priorizador para campañas comerciais que permita determinar a que clientes deben dirixirse as campañas de cada produto financeiro, cales son as canles de marketing relacionadas con cada cliente para realizar contratacións e, en consecuencia, a través de que canles deben realizarse as

comunicacións, tendo en conta as distintas restricións de negocio. Este proxecto desenvolve un modelo predictivo con diferentes métodos de aprendizaxe estatística para determinar cal é a canle preferente para as comunicacións comerciais e trata o problema inicial resolvendo un problema de optimización usando unha implementación do algoritmo de ramificación e corte. Este modelo de optimización incorpora as estimacións do modelo predictivo anterior e dos modelos de propensión á contratación de produtos financeiros xa dispoñibles na entidade xunto coas restricións inducidas polas necesidades empresariais.

Optimización do transporte de mostras biomédicas

Ángel Martínez González

Directora: Balbina Virginia Casas Méndez. USC. Modalidade académica

Revisión e aplicación do problema de optimización do transporte de mostras biomédicas. Propoñeremos unha descrición do estado da arte correspondente a este problema para realizar unha selección de modelos e abordar un caso práctico tomado da vida real. Interesa avaliar o rendemento destes modelos e concluímos o tamaño dos problemas que poden ser resoltos de maneira exacta ou aproximada e analizar os tempos de computación segundo o método usado. Ademais, consideraremos o deseño e implementación dun algoritmo heurístico que permita a resolución de problemas de maior tamaño.

Clasificación das nubes de puntos tridimensionais de parcelas forestais mediante técnicas de aprendizaxe profunda

Diego Laíño Rebollido.
Director: Javier Roca Pardiñas
Titor: Carlos Cabo Gómez (Uniovi)
UDC. Modalidade prácticas

Neste traballo propónse un modelo baseado nunha rede neuronal de perceptrón multi-capas que clasifica as estruturas presentes en nubes de puntos tridimensionais de parcelas forestais en 4 categorías: “arbusto”, “herba”, “póla” e “tronco”. Para isto, as nubes de puntos sométese a un procedemento de voxelización que reduce a súa complexidade e, tras isto, extráense descritores xeométricos baseados na análise de compoñentes principais da proximidade dos puntos, que farán de preditores das categorías. O axuste realízase por validación cruzada a partir de 8 microparcelsas obtidas mediante unha mostraxe pseudo-aleatoria da nube completa. O tratamento estatístico dos datos realizouse combinando as linguaxes de programación R e Python. O obxectivo final deste traballo é desenvolver unha técnica que facilite o inventariado forestal e, así mesmo, a estimación de parámetros que permitan asesorarse do estado de conservación dos bosques, o seu potencial económico ou, simplemente, sobre a súa estrutura en certo instante de tempo.

Xestión de camas no Complexo Hospitalario Universitario de Santiago de Compostela (CHUS).

Mario Picáns Rey.
Directoras: María Isabel Borrajo García e Mercedes Conde Amboage.
Titor: Francisco Reyes Santiás (Fundación Instituto de Investigación Sanitaria de Santiago de Compostela, FIDIS).
USC. Modalidade prácticas

O Servizo de Admisión e Documentación Clínica do Complexo Hospitalario Universitario de Santiago é o servizo encargado da xestión de pacientes do hospital, cuxos

propósitos principais son ordenar os fluxos dos mesmos e tramitar a información clínica que xeran. Co obxectivo de proporcionar ferramentas de control de ocupación aos diferentes servizos do hospital, neste traballo aplicaremos distintos modelos de regresión sobre os datos proporcionados, que conteñen información sobre as/os pacientes que ingresan no período 2016-2020. A finalidade de empregar os modelos que axustaremos no presente traballo é, entre outras, captar a natureza das variables de interese, como poden ser os ingresos diarios nun servizo. Os obxectivos deste traballo son a realización dunha revisión bibliográfica dos métodos que se empregan habitualmente para a modelización deste tipo de variables resposta e a creación de modelos que nos permitan prever a taxa de ocupación dun servizo, ou a duración da estada dos ingresos.

Análise da capacidade dispersiva e filopatría das poboacións do xénero *Alosa* no noroeste da península ibérica a través de técnicas de aprendizaxe automático

Alejandro Pico Calvo
Directores: Rosa María Crujeiras Casais e Fernando Cobo Gradín.
USC. Modalidade académica

Alosa alosa (Linnaeus, 1758) e *Alosa fallax* (Lacépède, 1803) son dúas especies de peixes diádomos cuxas poboacións están en declive e a súa escaseza e falta de información sobre o seu ciclo biolóxico, convertéronse nun problema emerxente. A microquímica dos otólitos de peixes diádomos correlaciona fortemente do seu río natal. No estudo que presentamos analizouse a composición microquímica dos otólitos de 141 adultos, distribuídos entre ambas especies, capturados tanto no mar coma no medio fluvial e comparouse coa dos otólitos de 137 xuvenís, repartidos entre ambas especies procedentes de diversos ríos do noroeste peninsular e de Francia. Deste xeito, mediante esta correlación púidose estudar o carácter filopátrico dos exemplares, concretar o seu río de nacemento e establecer o seu rango de dispersión durante a fase mariña. Así, mediante a aplicación de metodoloxías de aprendizaxe automática, achegamos nova información sobre o rango dispersivo destas especies e as interaccións entre poboacións.

Uso de técnicas SPC para o estudo e control dos indicadores da COVID-19

Claudia Torviso Rodríguez.
Directores: Javier Tarrío Saavedra e Salvador Naya Fernández.
UDC. Modalidade académica

Neste traballo faise unha análise do uso de ferramentas

propias do Control Estatístico de Procesos para o estudo da evolución de enfermidades, como é o caso da COVID-19. Lévese a cabo unha revisión da literatura sobre o SPC e dos Gráficos de Control máis utilizados. Analízanse as diferenzas entre o SPC na industria e no ámbito da saúde pública, e estúdanse os gráficos e métodos específicos máis utilizados neste último. Ademais, analizáronse os novos métodos de estudo xurdidos a raíz

da irrupción do SARS-CoV-2 no mundo. Como caso de estudo, aplicáronse algunhas das técnicas descritas ao longo do traballo a un caso concreto que é o estudo da COVID-19 nas augas residuais na área metropolitana de A Coruña. Estes datos foron cedidos polo proxecto de COVIDBENS, do cal faise unha breve introdución ao traballo levado a cabo. Por último, aplícanse os gráficos de control específicos en epidemioloxía a estes datos.

Traballos Fin de Grao

Inferencia estatística con datos censurados

Álvaro Añón Dosil

Director: César A. Sánchez Sellero

Titulación: Grao en Matemáticas. USC.

Neste traballo estúdanse algunhas das ferramentas máis habituais na análise de datos con censura, enmarcadas dentro da rama da Estatística denominada Análise de supervivencia. Para comezar, introducíronse unha serie de conceptos, definicións e funcións básicas recorrentes ao longo do traballo, como a función de supervivencia ou a función de risco. A continuación, explicouse o método de estimación de modelos paramétricos mediante máxima verosimilitude con datos censurados. Ademais, facendo uso de R fíxose un estudo de simulación para analizar as propiedades dos estimadores paramétricos, dos intervalos de confianza asintóticos e por perfil de verosimilitude. Máis adiante, presentáronse os estimadores non paramétricos da distribución, primeiro o que se obtén coas táboas de vida e despois mediante o estimador de Kaplan-Meier. Para rematar, mostrouse unha proba gráfica de bondade de axuste dos modelos paramétricos baseada no estimador de Kaplan-Meier. Ilustráronse estes métodos cos datos pharmacoSmoking do paquete asaur.

Introdución á simulación

Martina Basdediós Tilve

Directora: María de los Ángeles Casares de Cal

Titulación: Grao en Matemáticas. USC.

A simulación estatística permítenos estudar sistemas do mundo real a partir de modelos que imiten as súas características e que poidan ser implementados nun ordenador. Neste traballo, faremos unha introdución a diferentes técnicas que se empregan en simulación. Comezaremos facendo un repaso dos conceptos básicos de probabilidade e estatística que imos necesitar coñecer. A continuación, falaremos sobre os números pseudoalea-

torios e sobre como se pode contrastar a súa validez para ser empregados en simulación. Veremos tamén algúns métodos que podemos utilizar para xerar valores de distintas variables aleatorias, que serán fundamentais para construír os modelos dos sistemas reais. Finalmente, presentaremos un par de exemplos nos que imos simular sistemas da vida real, poñendo en práctica as ideas que veremos ao longo do traballo.

Métodos de Clasificación con datos obtidos mediante LiDAR

Andrea Blanco Pérez

Directores: Manuel Febrero Bande e María

José Ginzo Villamayor

Titulación: Grao en Matemáticas. USC.

O considerable aumento da aparición e severidade de incendios forestais nas últimas décadas, supuxo o inicio de investigacións orientadas a evitar e reducir o seu impacto e aparecemento. Neste marco, xorde o modelo Prometheus como un capaz de resumir de forma representativa a distribución das masas forestais. Este proceso está avalado pola tecnoloxía LiDAR, que é capaz de obter puntos tridimensionais cunha gran precisión e facilidade para percorrer áreas amplas. Neste traballo propónse a revisión e comparación de algoritmos de clasificación aplicados ao etiquetado de masas forestais seguindo o modelo Prometheus. Deste xeito, presentaranse distintos métodos aplicables ao problema de estudo, indicando o seu marco teórico e características. Concretamente, estudárase o método de Bayes, KNN e Modelos Lineais Xeneralizados no contexto do anterior problema. Finalmente, analizaranse os resultados obtidos tras a súa aplicación sobre nubes de puntos reais, comparando a súa eficacia e rendemento.

Regresión xeneralizada aplicada

Alba Camino Enríquez

Directores: Jose Ameijeiras Alonso

Titulación: Grao en Matemáticas. USC.

Os modelos de regresión serven para explicar e modelar a relación que existe entre unha variable resposta e unha ou máis variables explicativas. Tomando como base o modelo de regresión lineal simple clásico presentaremos distintas extensións que permitan xeneralizar dito modelo. En concreto, expoñeremos dous modelos de regresión sobre unha variable resposta discreta: o modelo de regresión loxística e o modelo de Poisson. É dicir, estes modelos en lugar de presentar unha distribución continua como era a normal para o caso lineal, presentan distribucións discretas como é a de Bernoulli e distribución de Poisson respectivamente. O modelo de regresión loxística aplícase cando a nosa variable resposta categórica é dicotómica. Mentres que, o modelo de Poisson é común empregalo para datos de conteo. Ambos os dous modelos, serán empregados sobre diferentes bases de datos relacionadas co ámbito da saúde e poderemos tratar as mesmas cuestións que para os modelos lineais.

Un percorrido polas distribucións notables na Inferencia Estatística con aplicacións

Pablo Carballo Fraguas
Director: Wenceslao González Manteiga
Titulación: Grao en Matemáticas. USC.

As distribucións de probabilidade son unha das ferramentas fundamentais dentro do campo da Inferencia Estatística. O obxectivo que se persegue neste traballo é levar a cabo un percorrido por algunhas das distribucións de probabilidade unidimensionais máis notables, tanto continuas como discretas, estudándoas desde o punto de vista das súas aplicacións e amosando algunhas das relacións que hai entre elas a través de resultados como o Teorema Central do Límite. Ademais, faise unha pequena introdución á simulación de variables aleatorias e amósanse algún dos principais algoritmos incluíndo exemplos prácticos. Por último, para facer fincapé na idea de que as distribucións de probabilidade non son só unha construción analítica senón que moitos fenómenos da realidade se axustan a estas leis estúdase a bondade de axuste de dúas distribucións, unha continua e unha discreta, a dous conxuntos de datos non simulados.

Curvas ROC

Raquel María Castro Capelo
Directora: Mercedes Conde Amboage
Titulación: Grao en Matemáticas. USC.

A curva ROC é unha ferramenta estatística amplamente empregada no ámbito sanitario para avaliar a capacidade diagnóstica dunha proba médica á hora de clasificar a unha poboación en dous grupos: pacientes enfermos e pacientes sans. Neste traballo revísanse os principais conceptos relacionados coa curva ROC xunto cos seus ín-

dices resumo, destacando a área baixo a curva e o índice de Youden, que é importante no momento de seleccionar un punto de corte óptimo en función dos obxectivos a estudo. Tamén se inclúen outros métodos para seleccionar o devandito limiar. Ademais, presentaranse, de maneira xeral, os principais métodos estatísticos para estimar a curva ROC en función do coñecemento da distribución que segue a variable diagnóstico asociada a cada categoría de interese. O funcionamento dos estimadores, tanto paramétricos como non paramétricos, ilustrarase grazas a datos simulados e á análise dunha base de datos reais.

A estimación non paramétrica da densidade

Laura Cotos Fernández-Arruty
Directora: María Isabel Borrajo García
Titulación: Grao en Matemáticas. USC.

Neste traballo sobre a estimación non paramétrica da densidade centrámonos no estudo do estimador tipo núcleo da densidade, que é un dos métodos non paramétricos máis empregados da Estatística. Introducimos os conceptos esenciais para a súa construción, así como para poder detallar e derivar as súas principais características. A cuestión máis esencial para o cálculo do estimador tipo núcleo na práctica é a elección do denominado parámetro ventá, para a que existen numerosos métodos dos que nós estudaremos os tres máis empregados: o selector de Silverman, o de Sheather e Jones e o de validación cruzada. Realizamos un estudo de simulación no que analizamos o comportamento destes tres selectores con diversos tamaños e catro modelos teóricos diferentes, confrontándoos ademais co valor óptimo teórico en cada caso. Rematamos o traballo cunha aplicación a datos reais do ámbito biomédico relativos a diversas magnitudes físicas do músculo cardíaco.

Modelos gráficos e redes neuronais para a estimación da densidade en alta dimensión

Manuel de Prada Corral
Directora: Rosa María Crujeiras Casais
Titulación: Grao en Matemáticas. USC.

A evolución tecnolóxica demanda a análise estatística de distribucións complexas con alta dimensionalidade, habitualmente tratadas con métodos de aprendizaxe automática. Para realizar tarefas de mostraxe e de estimación de densidade, na aproximación clásica de modelos paramétricos, impóñense condicións que faciliten os cálculos, xeralmente realizados mediante técnicas de máxima verosimilitude. Neste traballo afóndase nos conceptos e metodoloxías máis utilizados para a estimación da densidade nun contexto de alta dimensionalidade. En primeiro lugar, revísanse as ideas clásicas da estimación por máxima verosimilitude de modelos paramétricos, así como

as limitacións da estimación tipo núcleo neste campo e a posibilidade de empregar metodoloxías semiparamétricas. En segundo lugar, estúdanse as distintas estratexias que se poden considerar para mellorar e simplificar o proceso de estimación, principalmente da man dos modelos gráficos. Preséntanse as súas vantaxes e limitacións a través de NADE (Neural Autoregressive Density Estimation) e outros exemplos, empregando simulacións e exemplos prácticos.

Regresión non paramétrica

Ana Domínguez Fernández
Director: Alberto Rodríguez Casal
Titulación: Grao en Matemáticas. USC.

A regresión trata de analizar como unha variable X , chamada independente, inflúe noutra variable Y , chamada dependente. Existen dous enfoques diferentes: a regresión paramétrica e a non paramétrica. Neste traballo estúdanse e compáranse ambos, ao estimar o modelo de regresión o enfoque non paramétrico é unha alternativa aos modelos clásicos de regresión lineal, xa que proporcionan moita flexibilidade na forma do modelo. Xa dende a perspectiva da regresión non paramétrica, o estudo céntrase no cálculo do estimador Nadaraya-Watson. O estimador dependerá dun parámetro, chamado ventá. Estudaremos a influencia do parámetro ventá para atopar a que minimize o erro cadrado medio do estimador para conseguir o máis eficiente. Por último, ilústrase a aplicación práctica dos métodos presentados, abordando a estimación non paramétrica da función de regresión dun conxunto de datos reais e, en particular, o cálculo da ventá óptima do estimador.

Métodos de aproximación de solucións na Teoría de Xogos

Sara Gómez Feás
Directores: Paula Saavedra Nieves, Alejandro Saavedra Nieves
Titulación: Grao en Matemáticas. USC.

A teoría de xogos cooperativa céntrase principalmente en analizar os seguintes aspectos dunha interacción multi-axente: que coalicións serán formadas e como será repartido o beneficio xerado entre os individuos involucrados. Para realizar esta distribución seguindo criterios como a xustiza ou a equidade, establécense as chamadas regras de reparto, que serán consideradas solucións dos xogos cooperativos. A complexidade computacional do cálculo destas solucións aumenta de maneira considerable cando o número de axentes involucrados é relativamente grande, polo que é necesaria a súa aproximación. Ao longo deste traballo, repásanse os procesos de estimación baseados en técnicas de mostraxe presentados na literatura para o valor de Shapley e o valor de Banzhaf, así como

para as súas extensións a situacións con estruturas de prioridade, o valor de Owen e o valor de Banzhaf-Owen. Ademais, móstrase a implementación en R destes algoritmos e a súa aplicación ao conflito da cota de leite en Galicia.

Métodos Estatísticos aplicados ao Deporte

Daniela Gómez González
Director: Manuel Febrero Bande
Titulación: Grao en Matemáticas. USC.

A Estatística está presente no día a día, incluíndo deportes. Grazas a esta disciplina, ligas como a NBA xeran grandes cantidades de beneficios ao ano e optimizan diferentes facetas do xogo. O traballo trata de estudar o salario dos xogadores da liga de baloncesto americana na tempada 2020-2021, en función da súa actuación durante a tempada. O obxectivo é o de, mediante a teoría dos modelos de regresión, aplicar métodos lineais e introducir métodos non paramétricos buscando unha aplicación práctica e interpretable dos datos recollidos, co propósito de ver que características son as máis influentes á hora de predicir os soldos. Ademais estúdase se existen casos que supoñan irregularidades na proposta destes modelos e as súas posibles interpretacións analizando xogadores que son pagados por enriba ou por debaixo do seu soldo estimado.

Xogos cooperativos e a súa aplicación a problemas de tarificación

Marcos Gómez Rodríguez
Directoras: Balbina Virginia Casas Méndez, Laura Davila Pena
Titulación: Grao en Matemáticas USC.

Unha das aplicacións da teoría de xogos cooperativos atópase na determinación de tarifas aplicables aos usuarios dunha determinada instalación, cuxos custos de construción ou mantemento quérense recuperar, e onde certos criterios de eficiencia e de equidade inspiran a definición das solucións consideradas. Neste traballo, estúdase como repartir os custos fixos dunha autoestrada entre os seus usuarios a través de peaxes, usando xogos cooperativos con utilidade transferible, e propónse o uso de unións a priori para representar grupos, como os viaxeiros frecuentes ou camioneiros, que teñan un maior poder na negociación, conseguindo rebaixas nas súas tarifas. Para iso, demostráranse resultados relativos aos valores de Owen e Tijs coalicional, e introducíranse novos valores: o valor de Shapley-Tijs, e dous valores ponderados inspirados nos valores de Shapley e Tijs, respectivamente. Estes resultados serán aplicados a un caso real, concretamente, a datos de tráfico dunha sección da autoestrada AP-9.

Inferencia Estatística en procesos puntuais sobre grafos lineais

Ignacio González Pérez
Directores: Wenceslao González Manteiga,
María Isabel Borrajo García
Titulación: Grao en Matemáticas. USC.

Este traballo constitúe un percorrido que, dende o nivel dun graduado en Matemáticas, permite abordar problemas de inferencia non paramétrica sobre a función de intensidade de procesos puntuais en grafos lineais. Comézase coa estimación da función de densidade, focalizándose nos métodos tipo núcleo. De seguido introdúcese os procesos puntuais no plano euclidiano, para logo aumentar a complexidade e presentalos sobre grafos lineais, estudando en profundidade técnicas de estimación non paramétrica da función de intensidade. Neste punto introdúcese a comparación de dúas poboacións no marco dos procesos puntuais en grafos lineais. Preséntanse solucións innovadoras consistentes en tres estatísticos para contrastar a proporcionalidade de funcións de intensidade de dous procesos puntuais nun mesmo grafo lineal. Amósanse os resultados do exhaustivo estudo de simulación levado a cabo para comprobar o bo comportamento dos métodos propostos, e finalmente aplícanse a un conxunto de datos reais de accidentes de tráfico en Río de Janeiro.

Introdución aos Modelos Mixtos

Diego Losada González
Directora: Mercedes Conde Amboage
Titulación: Grao en Matemáticas. USC.

Moitas bases de datos concernentes a eidos como a Educación están xerarquicamente organizadas debido á propia natureza destas, de xeito que os individuos se atopan aninados en grupos. As/os alumnas/os clasificadas/os nunha mesma escola tenderán a ter un comportamento máis semellante que alumnado de escolas diferentes, e precísase ter en conta o efecto que producen estas agrupacións en certas variables de interese, empregando modelos clásicos como o ANOVA ou o ANCOVA. Afondando aínda máis, os grupos presentes neste tipo de conxunto de datos poden considerarse unha mostra aleatoria dunha poboación máis grande de grupos. É aquí onde xorden os modelos mixtos ou multinivel. Neste traballo, empréganse este tipo de modelos para estudar a nota acadada en Matemáticas por diferentes estudantes de distintas escolas. A incorporación de efectos aleatorios tanto en variables relativas ás/aos estudantes como ás escolas ás que pertencen resultan fundamentais para poder describir o comportamento da nosa variable de

interese.

Introdución ás redes neuronais

Ana Martínez Leboráns
Directoras: Balbina Virginia Casas Méndez,
Laura Davila Pena
Titulación: Grao en Matemáticas. USC.

A aprendizaxe profunda é unha disciplina da área da intelixencia artificial que está en auxe dada a súa adecuación para resolver problemas complexos. Este traballo recolle a historia da aprendizaxe profunda dende os seus inicios, así como os conceptos básicos necesarios para a súa comprensión. Introdúcense as definicións formais dalgunhas funcións de activación, distintos tipos de redes neuronais como as redes recorrentes ou convolucionais e algoritmos e conceptos relativos ao seu adestramento, entre os que se inclúen as definicións de varias funcións de custo e perda e métodos de optimización de algoritmos de aprendizaxe. Finalmente, amósanse algunhas das aplicacións que ten no mundo real.

Fiabilidade industrial en rodamentos de bólas presentes en dispositivos biomédicos

M^a Carmen Pousa Besada
Director: Juan Carlos Pardo Fernández
Titulación: Grao en Enxeñería Biomédica. UVigo.

Neste traballo aplícanse técnicas estatísticas para analizar a fiabilidade de dispositivos biomédicos. En concreto, trabállase con tempos de vida dun conxunto de 46 rodamentos de bólas, elementos presentes en dispositivos tan variados como ventiladores mecánicos, cadeiras de brazos odontolóxicas, TACs e brazos articulados, entre outros.

No traballo revísanse os conceptos estatísticos básicos necesarios para a análise da mostra, así como as distribucións máis empregadas en fiabilidade industrial, entre as cales a Weibull xoga un papel destacado. Introdúcese tamén o concepto de censura, xa que a principal dificultade que pode xurdir nun experimento de tempos de vida é que non todos os compoñentes fallen antes da finalización do estudo.

Os estudos de fiabilidade sobre os rodamentos permitirían realizar accións de mantemento preventivo mediante a súa substitución de forma previa ao fallo. Isto reduciría os custos derivados dun mal funcionamento ou da parada dos dispositivos, redundando na calidade da atención ao paciente.

Introdución aos modelos de inventario**Noelia Sánchez Martínez****Directora: María de los Ángeles Casares de Cal****Titulación: Grao en Matemáticas. USC.**

A motivación do traballo é a introdución das nocións básicas relativas á teoría de inventarios. Tras unhas breves notas históricas, estudarase a formulación, xustificación e utilidade dos principais modelos determinísticos así como as súas limitacións. No ámbito dos modelos estocásticos trátanse dous modelos dun só período: o modelo para produtos perecedoiros e o modelo de vendedor de xornais. Con respecto á calidade de servizo, céntrase no control da probabilidade de cumprir coa demanda mediante a utilización dos inventarios de seguridade, mostrando como estes se ven afectados polos sistemas de revisión continua e periódica de inventario e as políticas (R, Q) e (s, S) . Conclúese cun modelo de múltiples chanzos para a produción en serie, ilustrando a necesidade de coordinación entre as distintas instalacións implicadas.

Regras Discriminantes na Análise Multivariante e a súa adaptación á alta dimensión**Paula Soto Rodríguez****Director: Wenceslao González Manteiga****Titulación: Grao en Matemáticas. USC.**

Este traballo trata sobre a análise discriminante e a introdución dalgúns métodos de clasificación, tanto tradicionais como de creación máis moderna. No primeiro capítulo preséntanse as nocións básicas de clasificación, así como dúas regras discriminantes clásicas: a lineal e a cadrática. Ademais, descríbense algunhas xeneralizacións destas regras, como as máquinas de vector soporte, que buscan suplir moitas das deficiencias dos métodos máis tradicionais. No segundo capítulo explícanse diversas técnicas de clasificación adaptadas ao contexto do Big Data, onde a dimensión ou, noutras palabras, o número de variables, é maior que o tamaño da mostra. Finalmente, o último capítulo ilustra o funcionamento destas regras en conxuntos de datos simulados e mediante a súa aplicación a dúas bases de datos reais.

Se dirixes/titorizas traballos académicos no eido da Estatística e da Investigación de Operacións e queres que se publique o seu resumo no **informest**, contacta connosco.

Coñecendo a un socio:

Jacobo de Uña Álvarez

Jacobo de Uña Álvarez (A Coruña, 1972) é Catedrático de Universidade no Departamento de Estatística e Investigación Operativa da Universidade de Vigo, e investigador do CINBIO. Matemático de formación, coordina dende o ano 1998 o grupo *Inferencia Estatística, Decisión e Investigación Operativa (SiDOR)* en dita universidade.



Foto: CINBIO

Descríbenos o teu traballo

Investigo en métodos estatísticos non paramétricos, que non asumen ningunha estrutura sobre a poboación de interese. A nivel de aplicacións, centreime na Análise de Supervivencia, moi relevante en ensaios clínicos, epidemioloxía, fiabilidade industrial ou economía. Máis recentemente investiguei en métodos de comparacións múltiples, fundamentalmente para os problemas de alta dimensión que presentan os datos ómicos. Ademais, imparto varias materias da área a nivel de máster e doutoramento.

Cantos anos levas traballando en estatística ou I.O.?

Dende 1996. Máis dun cuarto de século!

Dende cando es socio da SGAPEIO?

Dende o ano 1993.

Cal foi o teu 1º Congreso da SGAPEIO?

O primeiro. Foi en A Coruña, en novembro do 1993.

O tema do teu próximo proxecto ou artigo?

Contrastes de bondade de axuste para modelos de regresión con resposta dobremente truncada.

Un curso que che gustaría facer?

Un curso de cociña.

Que libro de estatística ou I.O. recomendarías?

Ao estudantado interesado en Análise de Supervivencia recoméndolle o monográfico de Klein y Moeschberger (Springer). Para unha reciclaxe sobre estatística matemática, recomendaría o libro de Robert Serfling *Approximation Theorems of Mathematical Statistics* (Wiley).

Se non traballases no ámbito da estatística ou da I.O., que che gustaría ser?

Futbolista. Escritor. Músico. Piloto de Fórmula 1. Necesitaría varias vidas.

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

É unha palabra bastante fea que, etimoloxicamente, pouco ten que ver co que quere significar -os estatísticos non somos homes de estado. Habería que cambiala, pero *Ciencia de Datos* é aínda peor. Actualmente, a palabra *estadística* suxíreme a avaliación do rendemento deportivo (número de balóns perdidos), ou a descripción dunha economía (as estatísticas do paro). *Matemática* si que é unha palabra preciosa.

Que é o que máis valoras do labor da SGAPEIO?

Que somos nós, os galegos, xuntos facendo algo importante. Tendemos a pensar que, para triunfar, hai que irse lonxe -e quedarse alí-, cando non hai nada máis valioso que impulsar a cultura e o avance do coñecemento entre os teus, a todos os niveis, dende a educación primaria ata a universitaria, e despois desta, xa no eido profesional. A SGAPEIO foi formada por nós e para nós, sen deixar de estar aberta ao mundo, e iso faina enorme.

En que cres que deberíamos mellorar?

A SGAPEIO funciona moi ben. Quizais poderían organizarse máis actividades formativas para os máis novos: estudantes, profesores e investigadores na súa primeira etapa profesional. Ou incrementar a presenza en redes sociais, buscando unha imaxe máis atractiva, dinámica e moderna da Sociedade. Pero non cambiaría nada do esencial.

O censo de Stalin

por Covadonga Rodríguez-Moldes Rey

Os censos son unha fonte inesgotable de noticias, cada un coa súa intrahistoria. Nesta ocasión trataremos o Censo de Stalin de 1937, nun escenario de triste actualidade pola guerra entre Rusia e Ucraína, que foi analizado no libro de Argemino Barro *Una historia de Rus. Crónica de la guerra en el este de Ucrania*, un ensaio publicado no 2019. Na páxina 171, pódese ler:

“En 1937, o Goberno dispúxose a facer o primeiro censo de poboación en máis dunha década. Os datos preliminares proporcionaron un número sorprendente: a URSS tiña oito millóns de habitantes menos do esperado. Os catro altos responsables da Comisión Estatística, encargados de elaborar o censo, foron detidos e fusilados días despois. O censo xamais foi publicado.”

A situación en Rusia previa a ese censo era a seguinte: en 1929 Stalin quería facer de Rusia unha nación industrializada que avanzase en riqueza e poboación. Precisaba desenvolver unha industria pesada comprando maquinaria estranxeira, e dispor dunha clase obreira que defendese as ideas do marxismo-leninismo. Para isto necesitaba diñeiro que pensaba obter exportando o trigo producido por 25 millóns de pequenos propietarios agrarios, unha parte importante deles en Ucraína, por entón unha das repúblicas da URSS. Stalin decidiu poñer estas propiedades e o control da produción baixo a tutela do goberno. O seu plan púxose en marcha en 1935. Stalin estaba tan optimista que afirmou que a poboación aumentaba a razón de tres millóns de persoas ao ano e que en 1937 habería 180 millóns de “almas” na URSS. Nese ano organízase o censo.

Stalin implicouse persoalmente, nomeando unha comisión na que había nove políticos de moi alto rango e os membros da Comisión Estatística (a organización formal que realizaría el censo) e deseñou os discutibles formularios. O 6 de xaneiro de 1937 contouse a poboación en cidades, pobos, vilas, campos de traballos forzados do sistema Gulag, gardas de lugares fronteirizos, militares do Exército Roxo, persoal ferroviario e viaxeiros en tren.

O resultado preliminar do censo foi de 162 millóns de habitantes, moi lonxe das cifras oficiais e das que esperaba Stalin, e iso trouxo consecuencias inmediatas. O Comité Central do PC decretou que o censo fora organizado “de forma incorrecta e pouco profesional”, que era “unha grave violación dos fundamentos básicos da ciencia estatística”. O director do Instituto Soviético de Estatística foi arrestado e fusilado en setembro de 1937 ao igual que os seus colegas. Centenares de funcionarios foron despedidos e moitos executados, en especial en Ucraína e Kazaxistán. O censo foi declarado inválido e xamais publicado.

Unha das causas polas que o número de habitantes que mostrou o censo non foi o esperado foi o xenocidio ucraíno, a terrible gran fame coñecida como Holodomor (fame de masas) derivada da nova política agraria de Stalin nos anos 1932-33. Estímase que só en Ucraína houbo máis de 4 millóns de mortos. Deste censo debemos pensar na necesidade de que a ciencia, e neste caso, a estatística, se manteña independente dos poderes políticos... unha utopía?

informest é unha publicación cuatrimestral da SGapeio. Podes atopar os números anteriores xa publicados en www.sgapeio.es/, no apartado PUBLICACIONES.

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.

Equipo editorial:

María Gómez Rúa e Alejandro Saavedra Nieves.
Leticia Lorenzo Picado, María Martín Vila,
M^a Dolores Rico Ramos, Covadonga Rodríguez-
Moldes Rey, Olga Simón Cruz,
César Andrés Sánchez Sellero,
Javier Tarrío Saavedra e Antonio Vidal Vidal.

Revisión lingüística:

Trinidad Pazos Celis.

Coa colaboración de:

Germán Aneiros, Graciela Estévez, Ana Clavería,
Javier Roca, Brais González Rodríguez,
Jacobó de Uña Álvarez e Esther López Vizcaíno.

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

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