informest

Boletín da SGAPEIO. Novembro de 2026. Número 74

Colaboración
Software
Actividades da SGAPEIO
Novas
Novas do IGE
Traballos de Estatística e IO
Coñecéndonos
Sabías que?

Dirección:
Laura Davila Pena
Marta Sestelo Pérez

Comunicación coa SGAPEIO:
www.sgapeio.es
secretaria@sgapeio.es
boletin@sgapeio.es
@sgapeio
sgapeio
sgapeio
Sgapeio

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

Editorial

Laura Davila Pena – Vicepresidenta de Investigación Operativa

Escribo esta primeira editorial con ledicia e co compromiso que supón para min ter asumido a Vicepresidencia de Investigación Operativa da SGAPEIO. Agardo que durante estes anos poida compartir neste boletín as boas novas da nosa comunidade e celebrar os logros de compañeiras e compañeiros!

E, de feito, así é como comezo: celebrando logros. A SGAPEIO participou exitosamente cunha sesión invitada no *XLII Congreso Nacional de Estadística, Investigación Operativa y Ciencia de Datos* (SEIO 2026), celebrado en Santiago de Compostela o pasado setembro. Aproveito para felicitar o comité organizador e, en particular, a quen o liderou, María José Ginzo Villamayor e Alejandro Saavedra Nieves, dúas persoas que dedicaron moitos anos a esta Sociedade como membros do seu Consello Executivo e que seguen colaborando activamente con ela. Parabéns e grazas, Álex e María!

Neste número destacamos tamén o VI PhDay-EIO, o encontro do alumnado do programa de doutoramento en Estatística e Investigación Operativa, que se celebrou o pasado mes na Universidade de Vigo. Desde a SGAPEIO queremos felicitar a Héctor González Vázquez por ter recibido o premio á mellor comunicación oral que outorga a SGAPEIO, e que ademais nos achega neste boletín un artigo de divulgación. Agradecemos tamén ao comité organizador do VI PhDay-EIO o seu traballo e compromiso coa formación das novas xeracións de investigadoras e investigadores en Estatística e Investigación Operativa. É un pracer ver como a mocidade se xunta para contar e compartir ciencia. Que haxa moitas edicións máis do PhDay!

Felicitamos tamén á nosa compañeira Rosa M. Crujeiras Casais, que vén de recibir a Medalla de Investigación da Real Academia Galega de Ciencias na categoría Enrique Vidal Abascal, da sección de Matemáticas, Física e Ciencias da Computación. Rosa, este non é máis que un de tantos merecidos recoñecementos!

Este número tamén ten moito de homenaxe. Wenceslao González Manteiga xubilouse o pasado mes de agosto e foi nomeado profesor emérito da USC, culminando unha longa e frutífera traxectoria que nunca deixaremos de eloxiar. Moitos parabéns, Wences!

Cómpre destacar igualmente o percorrido exemplar da nosa compañeira Balbina Virginia Casas Méndez, que comeza unha nova etapa tras a súa xubilación. Balbina, a miña directora de tese: que alegría e, ao mesmo tempo, que pena me dá falar aquí desta despedida académica. Moitas grazas por estar sempre para min. Estou segura de que falo en nome de toda a SGAPEIO cando digo que fuches, es e seguirás sendo un piar fundamental da nosa Sociedade.

Despídome agradecendo a todas as persoas que, dunha forma ou doutra, colaborades co INFORMEST. Grazas polas vosas contribucións e por facer que esta Sociedade continúe medrando e fortalecéndose!

Con decembro xa á volta da esquina, aproveito para desexarvos un moi bo final de ano e un mellor comezo do seguinte. Vémonos na Asemblea do vindeiro día 22 e, se non, ata 2027!

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

Cando o obxecto a estimar é un conxunto

por Héctor González Vázquez

Héctor González Vázquez é investigador predoutoral da Xunta de Galicia en Estatística e Investigación Operativa na Universidade de Santiago de Compostela. É membro do Grupo MODESTYA e persoal investigador en formación do CITMAGA.

En estatística, cando falamos de estimar, habitualmente referímonos a aproximar un parámetro como un número ou un vector de dimensión finita. Así, por exemplo, buscamos estimar a media dunha distribución ou os coeficientes dun modelo de regresión linear. Tamén abordamos problemas nos que queremos estimar un obxecto de dimensión infinita, como na ben coñecida estimación non paramétrica de curvas, onde se pretende estimar funcións como a densidade ou a curva de regresión.

Cando o obxecto descoñecido é un subconxunto $S \subset \mathbb{R}^D$ e o obxectivo é reconstruílo a partir dunha mostra de puntos, entramos na área da estatística coñecida como estimación de conxuntos. O caso posiblemente máis paradigmático é a estimación do soporte: dispoñemos dunha mostra aleatoria $\mathcal{X}_n = \{X_1, \ldots, X_n\} \subset \mathbb{R}^D$ que procede dunha certa distribución de probabilidade e o noso obxectivo é estimar o seu soporte, que é o menor conxunto pechado con probabilidade un. Intuitivamente, interésanos estimar o soporte cando queremos coñecer a rexión de onde proveñen os datos. Algúns contextos nos que aparece este problema inclúen a reconstrución da forma dun obxecto nunha imaxe ou a estimación do hábitat dunha especie a partir de observacións de exemplares da mesma.

Gatos pescadores nunha paisaxe transformada polo ser humano

O gato pescador (*Prionailurus viverrinus*) é un tipo de gato salvaxe que vive maiormente preto das zonas húmidas (humidais) no sur e sueste de Asia e que recibe ese nome porque se alimenta principalmente de peixes. Está catalogado como especie vulnerable na Lista Vermella da Unión Internacional para a Conservación da Natureza (UICN), xa que se ve ameazado pola desaparición das zonas húmidas. A zona no entorno do parque nacional de Khao Sam Roi Yot, na costa occidental do golfo de Tailandia, ilustra ben este problema. O parque protexe unha marisma de auga doce que está recoñecida como zona húmida de importancia internacional, pero arredor dela a paisaxe modificouse ao longo dos anos pola actividade humana, converténdoa nunha zona con aldeas, plantacións e estanques de gambas ou peixes, moitos deles despois abandonados.

Neste contexto, resulta natural preguntarse que áreas utilizan realmente os gatos e estudar como o seu hábitat se adapta a estas transformacións humanas sobre as zonas húmidas. Para abordar esta cuestión, un grupo de investigadores recolleu datos GPS de 13 gatos pescadores (Khaewphakdee et al., 2025). Os datos, dispoñibles a través de Movebank, unha plataforma internacional para o arquivo e seguimento de datos de movemento animal, proceden de colares que rexistraban a posición dos animais en instantes de tempo consecutivos, cunha separación entre medicións que variaba de 10 minutos a 4 horas, e nun rango de semanas ou meses distinto para cada gato que en total abarcou dende xuño de 2022 a marzo de 2024; véxase a Figura 1.

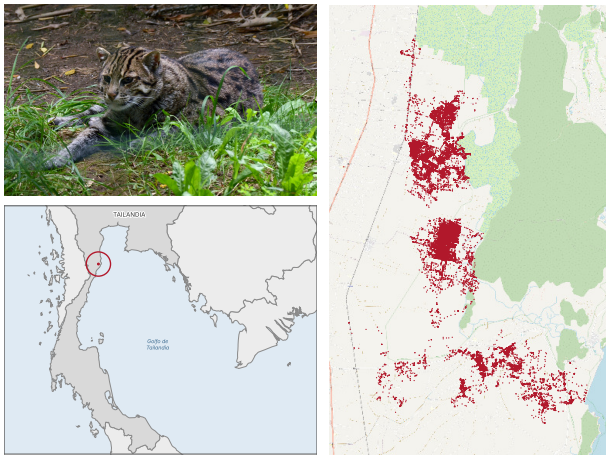

Figura 1: Arriba á esquerda, gato pescador (*Prionailurus viverrinus*). Abaixo á esquerda, localización do parque nacional de Khao Sam Roi Yot en Tailandia. Á dereita, nube de puntos coas posicións GPS dos gatos do estudo.

A partir destas localizacións pódese formular un problema típico de estimación do soporte: utilizando só os puntos observados, queremos reconstruír a rexión do espazo na que se moven os animais. En ecoloxía, o *home range* ou área de vida dun animal é a zona que percorre nas súas actividades normais, como a busca de alimento, apareamento e cría. Neste caso, a estimación do soporte permite obter unha representación xeométrica das zonas utilizadas polos gatos e comparala coas características da paisaxe. De feito, os investigadores observaron que os gatos prefiren os estanques, especialmente aqueles abandonados con maior vexetación de marisma, mentres que evitan os cultivos e as zonas poboadas (Klakhaeng et al., 2024; Serieys et al., 2025).

Como estimar o soporte? A envoltura convexa

Un dos estimadores máis sinxelos e que primeiro se consideraron na literatura de estimación do soporte é a envoltura convexa da mostra (Rényi e Sulanke, 1963), isto é, o menor conxunto convexo que contén as observacións:

$$\hat{S}_n = \text{conv}(\mathcal{X}_n).$$

En realidade, ten unha longa historia en ecoloxía, aínda que con outra denominación pois, baixo o nome de mínimo polígono convexo, a envoltura convexa leva moitas décadas sendo o estimador de referencia para o *home range* dende que foi proposto por Mohr (1947). O motivo radica en que é un estimador sinxelo de comprender e empregar, xa que non depende de ningún parámetro, así como de representar, pois no plano é un polígono que ten como vértices algunhas das observacións mostrais.

Na Figura 2 represéntase a envoltura convexa das posicións dos gatos pescadores. Estimaría ben o soporte que queremos recuperar baixo a suposición de que este sexa convexo. Porén, á vista dos nosos datos, este non parece ser o caso, pois o estimador enche unha gran cantidade de área na que non hai observacións. Por este motivo, precisamos estimadores capaces de capturar formas máis xerais.



Figura 2: Envoltura convexa.

Un estimador baseado no suavizado

Para obter un estimador non paramétrico do soporte podemos pensar en empregar un método de suavizado, cunha idea similar ao estimador tipo núcleo da densidade. Se un punto pertence á mostra, intuitivamente o seu entorno máis inmediato debe pertencer ao soporte. Deste xeito, podemos, para cada observación, colocar unha bóla pechada dun certo raio $\varepsilon > 0$ arredor dese punto, obtendo como estimador do soporte a seguinte unión de bólas

$$\hat{S}_n = \bigcup_{i=1}^{n} B(X_i, \varepsilon).$$

Este coñécese como o estimador de Devroye–Wise (Devroye e Wise, 1980). Na Figura 3 represéntase o estimador para valores crecentes de ε.

O valor do raio das bólas ε xoga o papel dun parámetro de suavizado. Se é demasiado pequeno, queda unha estimación infrasuavizada e o estimador pode chegar a ter demasiadas compoñentes conexas, algunhas delas formadas por unha única illa arredor dun punto. Pola contra, un raio demasiado grande devolve un estimador sobresuavizado, onde as bólas son tan grandes que se perde moito detalle da nube de puntos orixinal, ademais de que se sobreestima notablemente o soporte.

Para garantir a consistencia deste estimador de modo moi xeral, as condicións asintóticas que se impoñen sobre o raio $\varepsilon \equiv \varepsilon_n$ son moi similares ás que se impoñen sobre a ventá h_n na estimación non paramétrica da densidade: ε_n debe tender a cero, pero non demasiado rápido, basta con que $n\varepsilon_n^D \to \infty$, para garantir que dentro das bólas cae un número de observacións cada vez maior.

Xeneralizando a condición de convexidade

A gran vantaxe do estimador de Devroye–Wise é a súa xeneralidade: permite entrantes, buracos e distintas compoñentes conexas no soporte a estimar. Porén, o prezo a pagar radica na súa eficiencia, xa que non presenta unha velocidade de converxencia rápida. Podemos buscar un equilibrio cun estimador que empregue unha condición xeométrica pero que esta non sexa tan restritiva como a convexidade? Unha posible resposta está no concepto de r-convexidade. Imaxina que tomas unha goma de borrar cuxa forma é unha bola aberta de raio $r > 0$ e, partindo de todo o plano, vas borrando con esa goma todo o que poidas sen tocar ningún punto da mostra. O que queda sen borrar é a envoltura r-convexa da mostra. Formalmente, defínese este estimador como

$$\hat{S}_n = C_r(\mathcal{X}_n) = \bigcap_{\mathring{B}(x,r)\cap\mathcal{X}_n=\varnothing} \left(\mathbb{R}^D \setminus \mathring{B}(x,r)\right).$$

Neste caso non é o estimador, senón o seu complementario, o que é unha unión de (infinitas) bólas. Deste xeito, a súa fronteira está formada por arcos de circunferencias de raio r, como se aprecia na Figura 4. O estudo da envoltura r-convexa como estimador do soporte foi abordado por Rodríguez-Casal (2007), mentres que a súa implementación no plano pódese atopar no paquete `alphahull` de R (Pateiro-López e Rodríguez-Casal, 2010).

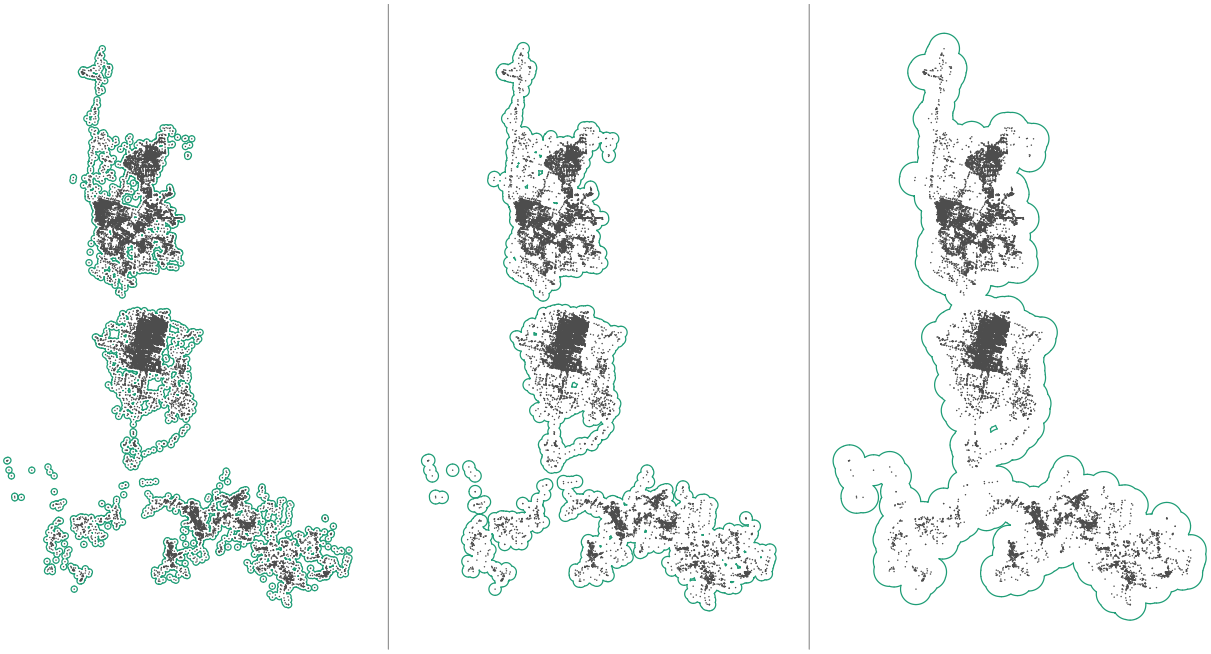

Figura 3: De esquerda a dereita, estimador de Devroye–Wise para $\varepsilon = 100$, 200 e 500 m.

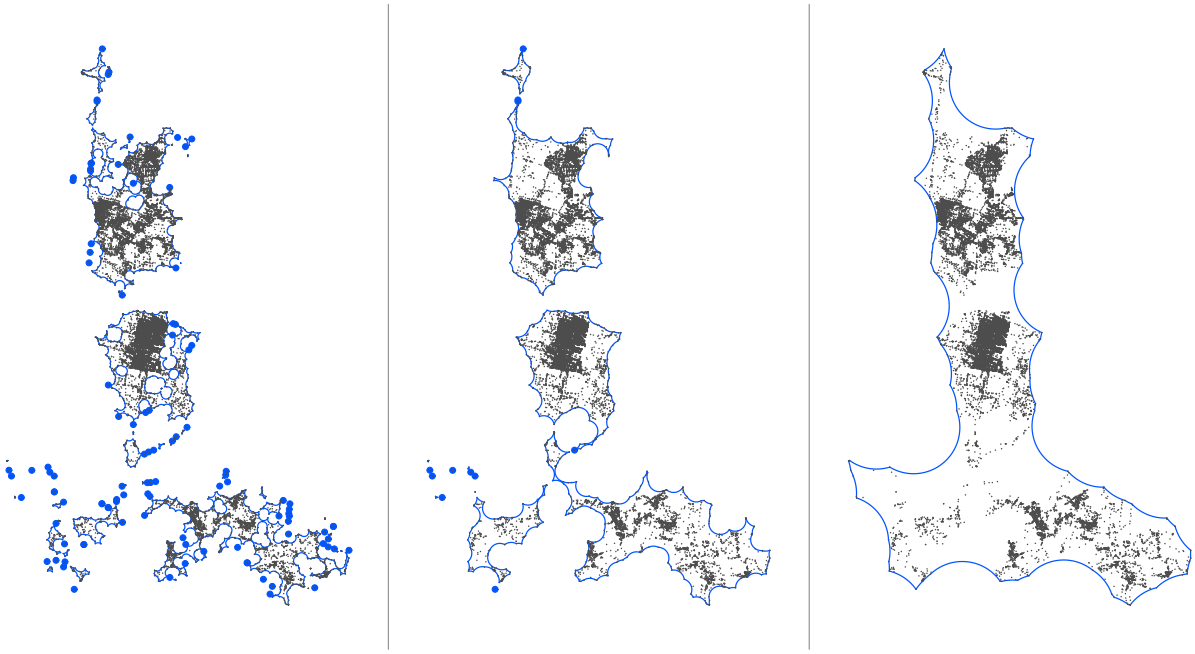

Figura 4: De esquerda a dereita, envoltura r-convexa para $r = 150$, 350 e 1500 m.

O parámetro r xoga o papel dun parámetro de forma. Hai unha monotonía do estimador respecto de r; valores moi pequenos devolven unha envoltura moi fragmentada, mentres que para valores grandes o estimador tende a encher ocos igual que facía a envoltura convexa. De feito, a envoltura convexa, cuxo complementario é unha unión de semiplanos, pódese ver como o límite cando $r \to \infty$ da envoltura r-convexa, xa que ao aumentar o raio as bólas parécense cada vez máis a semiplanos. Porén, a diferenza do ε do Devroye–Wise, r non é un parámetro de suavizado no sentido habitual do termo en estatística non paramétrica, xa que non debe tender a cero. Pódese ver máis ben como unha escala de resolución á cal observamos a nube de puntos.

Máis alá do soporte: outros problemas en estimación de conxuntos

A estimación do soporte é só a porta de entrada á estimación de conxuntos. En función do estudo que queiramos realizar, pode ser interesante reconstruír outros conxuntos ou incluso certas características dos mesmos.

Conxuntos de nivel e rexións de alta densidade. A estimación do soporte trata igual as zonas nas que só se rexistrou algunha posición dalgún gato que aquelas onde pasan a maior parte do tempo e nas que se observan máis puntos. Por este motivo, ás veces, en lugar de estimar todo o soporte, queremos estimar unha rexión de alta densidade. Formalmente, se definimos un conxunto de nivel como a rexión onde a densidade f sobrepasa un certo limiar λ,

$$L(\lambda) = \{x \in \mathbb{R}^D : f(x) \geq \lambda\},$$

unha rexión de alta densidade é un conxunto de nivel que contén unha probabilidade fixada $1-\alpha$, por exemplo do $90\,\%$. Ademais, os conxuntos de nivel relaciónanse coa análise de clústeres: as compoñentes conexas de $L(\lambda)$ pódense interpretar como grupos.

Funcionais do conxunto. Noutras ocasións, o conxunto en si non é o obxecto de estudo, senón algunha das súas características. En ecoloxía, a estimación do *home range* adoita buscar aproximar a superficie (por exemplo en km^2) que abarca a zona que percorre o animal, o que matematicamente correspondería coa súa medida de Lebesgue. Este problema non é tan sinxelo como parece: por exemplo, se se emprega un enfoque *plug-in* e se estima a área do soporte mediante a área da envoltura r-convexa, resulta que o estimador sofre dun nesgo que infraestima a verdadeira área, polo que é necesario considerar estimadores máis sofisticados (Arias-Castro et al., 2019).

Estruturas de dimensión inferior. Pode ocorrer que os datos realmente se distribúan arredor de estruturas de dimensión inferior; por exemplo, os gatos poderían desprazarse case sempre seguindo as liñas que trazan os diques que separan os estanques ou certos camiños. Neste caso, en lugar de reconstruír un conxunto bidimensional, o interese estaría en recuperar a estrutura das curvas que subxacen aos datos, que son unidimensionais. En dimensión superior, estas estruturas correspóndense co concepto matemático de variedade.

Algúns detalles que omitimos

Polo camiño deixamos fóra certas cuestións, algunhas delas máis técnicas. O lector interesado nunha panorámica máis ampla da estimación de conxuntos pode consultar Cuevas (2009); Cholaquidis (2024). Para unha revisión da estimación do *home range* dun animal véxase Baíllo e Chacón (2021).

Datos en variedades. As posicións dos gatos realmente non viven no plano euclidiano, senón que están na superficie da esfera. Nunha rexión dunhas poucas decenas de quilómetros cadrados a proxección ao plano apenas distorsiona a xeometría, pero se considerásemos unha especie con movementos migratorios ou cun hábitat repartido por todo o planeta si sería un problema. Por iso, cando traballamos con datos que viven nunha xeometría non euclidiana, é necesario empregar técnicas adaptadas a esta situación.

Dependencia nos datos. A maior parte da teoría de estimación de conxuntos asume que os datos proveñen dunha mostra aleatoria simple e polo tanto son independentes. A situación que presentan os datos dos gatos pescadores é moi distinta: hai unha forte dependencia entre a medición nun instante de tempo e a do instante posterior. Un posible enfoque alternativo sería tratar os datos como puntos observados densamente ao longo dunha traxectoria continua para estimar o soporte a partir destas traxectorias.

Como se mide o erro. Os nosos estimadores son conxuntos aleatorios e necesitamos empregar unha distancia entre conxuntos para falar de consistencia e de velocidades de converxencia. As dúas distancias máis habituais entre dous conxuntos son a distancia de Hausdorff, que mide canto hai que "engordar" cada conxunto para que cubra ao outro, e a distancia en medida, que representa a área que está nun dos dous conxuntos pero non no outro.

Condicións de forma. Para obter taxas de converxencia hai que fixar hipóteses sobre o soporte ou conxunto a estimar, do mesmo xeito que na estimación de densidades hai que supoñer certa suavidade. Estas hipóteses son condicións de forma do estilo da r-convexidade, que habitualmente teñen unha interpretación xeométrica intuitiva, o que fai da estimación de conxuntos unha área onde a xeometría xoga un rol esencial.

Agradecementos

O autor agradece aos profesores Beatriz Pateiro e Alberto Rodríguez todas as súas suxestións e a axuda brindada para mellorar o texto.

Os mapas da Figura 1 están realizados con OpenStreetMap. As Figuras 3 e 4 foron realizadas co paquete `alphahull` de R (Pateiro-López e Rodríguez-Casal, 2010).

O autor agradece o apoio económico da axuda de apoio á etapa predoutoral da Xunta de Galicia ED481A-2025-173, cofinanciada parcialmente pola Unión Europea no marco do programa FSE+ Galicia 2021-2027. Este traballo é parte do proxecto PID2024-158017NB-I00 financiado por MICIU/AEI/10.13039/501100011033 e por FEDER, UE, e o proxecto ED431C 2025/03, financiado pola Consellería de Educación, Ciencia, Universidades e Formación Profesional.

Referencias

Arias-Castro, E., Pateiro-López, B., & Rodríguez-Casal, A. (2019). Minimax estimation of the volume of a set under the rolling ball condition. *Journal of the American Statistical Association*, 114(527), 1162-1173.

Baíllo, A., & Chacón, J. E. (2021). Statistical outline of animal home ranges: An application of set estimation. En Srinivasa Rao, A. S. R., & Rao, C. R. (eds.), *Data Science: Theory and Applications*, vol. 44, pp. 3-37. Elsevier.

Cholaquidis, A. (2024). New advances in set estimation. *Boletín de Estadística e Investigación Operativa*, 40(1), 5-21.

Cuevas, A. (2009). Set estimation: another bridge between statistics and geometry. *Boletín de Estadística e Investigación Operativa*, 25(2), 71-85.

Devroye, L., & Wise, G. L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. *SIAM Journal on Applied Mathematics*, 38(3), 480-488.

Khaewphakdee, S., Serieys, L. E. K., & Pattanarangsan, R. (2025). Data from: Study "Fishing cat (Prionairulus viverrinus) study in Khao Sam Roi Yot, Thailand". Repositorio de datos Movebank. https://doi.org/10.5441/001/1.644

Klakhaeng, C., Khaewphakdee, S., Mongkonsin, W., Serieys, L. E., Wong, W.-M., Yindee, M., Pattanarangsan, R., Suksavate, W., Promkuntod, P., Wongson, T., & Sukmasuang, R. (2024). Home range and factors affecting the appearance of the fishing cat (Prionailurus viverrinus) in a human-dominated landscape, Thailand. *Journal of Wildlife and Biodiversity*, 8(4), 311-328.

Mohr, C. O. (1947). Table of Equivalent Populations of North American Small Mammals. *The American Midland Naturalist*, 37(1), 223-249.

Pateiro-López, B., & Rodríguez-Casal, A. (2010). Generalizing the convex hull of a sample: the R package alphahull. *Journal of Statistical Software*, 34(5), 1-28.

Rényi, A., & Sulanke, R. (1963). Über die konvexe Hülle von n zufällig gewählten Punkten. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 2(1), 75-84.

Rodríguez-Casal, A. (2007). Set estimation under convexity type assumptions. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, 43(6), 763-774.

Serieys, L. E. K., Khaewphakdee, S., Mongkonsin, W., Kaewplang, K., Yindee, M., Suksavate, W., Klakhaeng, C., Wongson, T., Sukmasuang, R., Wong, W.-M., & Pattanarangsan, R. (2025). Adapting to change: A wetland-specialist carnivore's habitat selection in a human-dominated wetland. *Biological Conservation*, 310, 111382.

`sicure`: Single-Index Mixture Cure Models

por Beatriz Piñeiro Lamas, Ana López Cheda e Ricardo Cao Abad

Introdución

Na análise de supervivencia, os *mixture cure models* aplícanse en situacións nas que unha fracción dos individuos nunca experimentará o evento de interese (individuos curados). Estes modelos permiten estimar tanto a probabilidade de curación como a función de supervivencia dos suxeitos susceptibles (latencia). Así mesmo, os *single-index mixture cure models* empréganse para tratar a maldición da dimensionalidade cando hai múltiples covariables presentes. O paquete `sicure` (Piñeiro-Lamas et al., 2025) inclúe funcións para axustar estes modelos dependendo de se a covariable é un vector, unha función ou unha combinación de ambos. Tamén proporciona un estimador non paramétrico para a función de densidade condicional da poboación susceptible.

Principais achegas do paquete

Na actualidade existen varios paquetes de R para modelos de curación, tanto paramétricos (`flexsurvcure`, `cuRe`, `geecure`) como semiparamétricos (`smcure`, `miCoPTCM`). No ámbito non paramétrico, `npcure` implementa estimadores para a probabilidade de curación e a latencia, aínda que limitados a covariables unidimensionais. Non obstante, ata o de agora existía un baleiro na dispoñibilidade de ferramentas para axustar modelos de curación con múltiples covariables mantendo unha compoñente non paramétrica que aporte flexibilidade ao modelo. O paquete `sicure` implementa os *single-index mixture cure models* propostos en Piñeiro-Lamas et al. (2026a,b). Especificamente, estende ao caso multivariante, mediante o enfoque *single-index*, os enfoques non paramétricos previos para covariables unidimensionais (Xu e Peng, 2014; López-Cheda et al., 2017a,b). Tamén xeneraliza ao caso con presenza de curación o modelo *single-index* para datos censurados pola dereita proposto en Strzalkowska-Kominiak e Cao (2013).

Uso do paquete

Notación

Na análise de supervivencia con datos censurados pola dereita consideramos as variables (T, δ), onde $T = \min(Y, C)$ é o tempo de supervivencia observado, Y o tempo de supervivencia real, C o tempo de censura e $\delta = I(Y \leq C)$ o indicador de non censura. Sexa $\mathbf{X} = (X_1, \ldots, X_d) \in \mathbb{R}^d$ un vector de covariables. Un *mixture cure model* expresa a función de supervivencia condicional como

$$S(t|\mathbf{x}) = 1 - p(\mathbf{x}) + p(\mathbf{x})\, S_0(t|\mathbf{x}),$$

onde $1 - p(\mathbf{x}) = P(Y = \infty|\mathbf{X} = \mathbf{x})$ é a probabilidade de curación e $S_0(t|\mathbf{x}) = P(Y > t|Y < \infty, \mathbf{X} = \mathbf{x})$ é a latencia. Se $d > 1$, os modelos *single-index* abordan a maldición da dimensionalidade resumindo o vector de covariables nun índice unidimensional mediante unha combinación lineal. Isto é, asúmese que existe un vector de parámetros $\boldsymbol{\theta_0} \in \mathbb{R}^d$ tal que

$$1 - p(\mathbf{x}) = 1 - p_{SI}(\boldsymbol{\theta_0}^\top \mathbf{x}) \quad \text{e}$$
$$S_0(t|\mathbf{x}) = S_{0,SI}(t|\boldsymbol{\theta_0}^\top \mathbf{x}),$$

onde $p_{SI} : \mathbb{R} \to [0,1]$ e $S_{0,SI} : \mathbb{R}^2 \to [0,1]$. Isto reduce a dimensión da covariable a un, mentres que se mantén a flexibilidade nas estimacións non paramétricas de $1-p_{SI}$ e $S_{0,SI}$. Polo tanto, os *single-index mixture cure models* poden definirse como:

$$S(t|\mathbf{X} = \mathbf{x}) = S_{SI}(t|\boldsymbol{\theta_0}^\top \mathbf{X} = \boldsymbol{\theta_0}^\top \mathbf{x}),$$

onde $S : \mathbb{R}^{d+1} \to [0,1]$ e $S_{SI} : \mathbb{R}^2 \to [0,1]$ é unha función tal que $S_{SI}(t|u) = 1 - p_{SI}(u) + p_{SI}(u) S_{0,SI}(t|u)$. O vector $\boldsymbol{\theta_0}$ estímase por máxima verosimilitude. A función de verosimilitude depende de $1 - p_{SI}$, $S_{0,SI}$ e da densidade condicional dos susceptibles, $f_{0,SI}$, que, ao ser descoñecidas na práctica, se estiman mediante estimadores tipo núcleo, o que introduce catro ventás.

Principais funcións

Ademais das tres funcións de axuste (`sicure.v`, `sicure.f` e `sicure.vf`), nas que nos centraremos, o paquete inclúe funcións auxiliares para estimar a probabilidade de curación de forma non paramétrica suavizada (`probcure.sm`), a densidade condicional da poboación susceptible por validación cruzada (`cd.uncured`, e a súa versión suavizada `cd.sm.uncured`), a log-verosimilitude do modelo (`loglik.simcm`) e a función obxectivo da optimización (`fun.opt`), así como o núcleo de Epanechnikov reescalado (`k.epa`) e os k veciños máis próximos coa distancia de Mahalanobis (`kNN.Mahalanobis`).

`sicure.v`

Esta función estima o vector de parámetros, $\boldsymbol{\theta_0} \in \mathbb{R}^d$, nun *single-index mixture cure model* con covariables vectoriais. Os argumentos son:

`x_cov`: matriz ou data frame cos valores das covariables. Cada fila representa un individuo e cada columna representa unha variable;

`time`: vector numérico cos tempos observados;

`delta`: vector numérico cos valores do indicador de non censura (1 se o evento foi observado, 0 se é censurado);

`multistart`: valor lóxico (`FALSE` por defecto) que especifica se se considera unha optimización con múltiples puntos de inicio.

A función devolve unha lista coas seguintes compoñentes:

`par`: vector numérico de dimensión $d+3$ coas estimacións dos parámetros. As primeiras $d-1$ compoñentes corresponden a $\hat{\boldsymbol{\theta}}_n$ (a primeira compoñente de $\boldsymbol{\theta_0}$ fíxase a 1 para evitar problemas de identificabilidade do modelo) e as 4 últimas aos logaritmos das ventás seleccionadas;

`value`: valor da función obxectivo (log-verosimilitude negativa) nos parámetros estimados;

`si`: variable *single-index* estimada.

`sicure.f`

Esta función axusta o modelo cando a covariable é funcional. Neste caso, `x_cov` contén os valores da función discretizada nunha grella de puntos, e a súa dimensión redúcese mediante Análise de Compoñentes Principais Funcionais (FPCA), considerando as K primeiras compoñentes necesarias para explicar unha proporción mínima da variabilidade, indicada polo argumento `propvar` (0.9 por defecto). O índice constrúese entón como unha combinación lineal das K puntuacións principais, con vector de coeficientes $\boldsymbol{\beta_0} \in \mathbb{R}^K$. Os demais argumentos e a saída son análogos aos de `sicure.v`: `par` ten agora dimensión $K+3$, e as súas primeiras $K-1$ compoñentes corresponden a $\hat{\boldsymbol{\beta}}_n$ (de novo, coa primeira compoñente de $\boldsymbol{\beta_0}$ fixada a 1).

`sicure.vf`

Esta función axusta o modelo cando se dispón simultaneamente de covariables vectoriais e dunha covariable funcional. O índice constrúese como unha combinación lineal das d covariables vectoriais e das K puntuacións principais obtidas mediante FPCA, con vector de coeficientes $\boldsymbol{\gamma_0} \in \mathbb{R}^{d+K}$. Os argumentos son similares aos de `sicure.v` e `sicure.f`. Na lista devolta, `par` ten dimensión $d+K+3$ e as súas primeiras $d+K-1$ compoñentes corresponden a $\hat{\gamma}_n$ (coa primeira compoñente de $\boldsymbol{\gamma_0}$ fixada a 1).

Aplicación a datos reais

Para presentar o uso do paquete `sicure`, consideramos un conxunto de datos que contén información sobre 531 mulleres diagnosticadas de cancro de mama, tratadas con fármacos potencialmente cardiotóxicos no Hospital Universitario da Coruña entre 2007 e 2021 (Piñeiro-Lamas et al., 2023a,b). O evento de interese é a aparición de cardiotoxicidade, un efecto secundario asociado ao tratamento, que desenvolveron 54 pacientes (10.17 %) durante o seguimento. Considéranse as variables continuas frecuencia cardíaca, idade, peso, estatura, LVEF, PWT, LAd, LVDd e LVSd. Como covariable funcional considérase a curva que describe a velocidade á que o corazón se contrae e se relaxa ao longo dun latexo, medida antes do inicio do tratamento, que informa do estado cardiovascular basal da paciente.

A modo ilustrativo, mostraremos os resultados da aplicación da función `sicure.vf`. No primeiro paso, obtemos que o número de compoñentes principais da covariable funcional que explican o 90 % da súa variabilidade é 8. As estimacións dos coeficientes para frecuencia cardíaca, idade, peso, estatura, LVEF, PWT, LAd, LVDd e LVSd son 1, 0.2307, 0.0976, 0.0306, -0.031, -0.2034, 0.0201, -1.1315 e 1.3864, respectivamente. A Figura 5 (esquerda) mostra a probabilidade estimada de sufrir cardiotoxicidade condicionada á variable *single-index*: canto maior é o valor desta variable, maior é a probabilidade de sufrir o problema cardíaco, o que indica que a variable *single-index* estimada podería utilizarse como factor de risco de cardiotoxicidade. Así, as covariables con coeficientes estimados positivos poden aumentar o risco, mentres que as de coeficientes negativos poden ter un efecto protector. Ademais, dado que as covariables foron estandarizadas, os coeficientes próximos a cero suxiren que as covariables correspondentes non son relevantes. Por exemplo, a idade da paciente aumentaría o risco de cardiotoxicidade, mentres que unha LVEF basal elevada protexeríaa fronte a ela, resultados coherentes coa literatura en cardio-oncoloxía. A Figura 5 (dereita) mostra a latencia, isto é, a función de supervivencia das pacientes que van sufrir cardiotoxicidade, para tres valores da variable *single-index* estimada (-1, 0 e 2): canto maior é o valor, antes se produce a aparición da cardiotoxicidade.

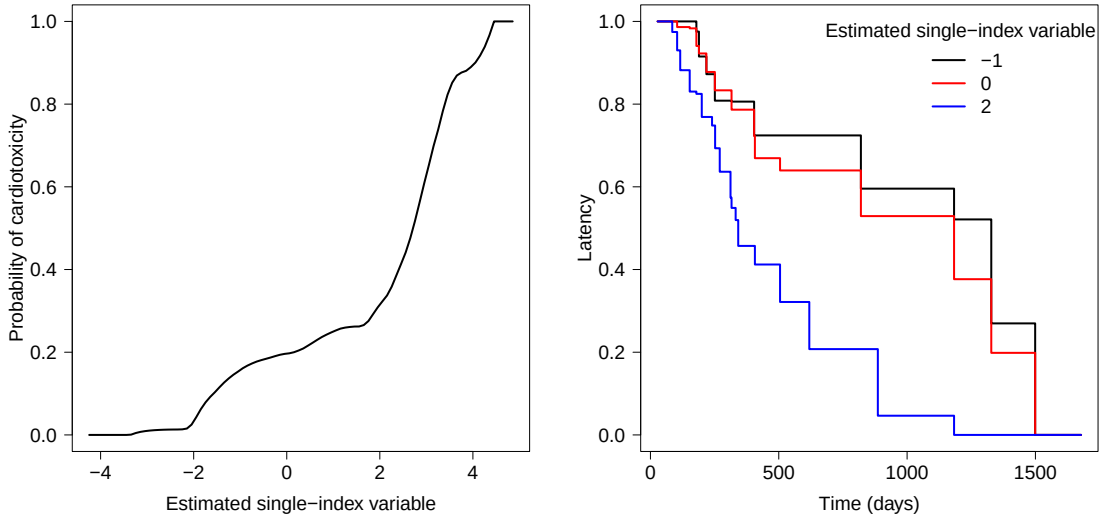

Figura 5: considerando os datos vectoriais e funcionais do conxunto de cardiotoxicidade, probabilidade de cardiotoxicidade estimada dependendo da variable tipo *single-index* (esquerda); e estimación da latencia para tres valores da variable *single-index* (dereita).

Referencias

López-Cheda, A., Cao, R., Jácome, M. A. e Van Keilegom, I. (2017a). Nonparametric incidence estimation and bootstrap bandwidth selection in mixture cure models. *Computational Statistics & Data Analysis*, 105, 144–165. https://doi.org/10.1016/j.csda.2016.08.002.

López-Cheda, A., Jácome, M. A. e Cao, R. (2017b). Nonparametric latency estimation for mixture cure models. *TEST*, 26, 353–376. https://doi.org/10.1007/s11749-016-0515-1.

Piñeiro-Lamas, B., López-Cheda, A., Cao, R., Ramos-Alonso, L., González-Barbeito, G., Barbeito-Caamaño, C. e Bouzas-Mosquera, A. (2023a). BC_cardiotox: A cardiotoxicity dataset for breast cancer patients. *figshare*. Dataset. https://doi.org/10.6084/m9.figshare.22650748.v4.

Piñeiro-Lamas, B., López-Cheda, A., Cao, R., Ramos-Alonso, L., González-Barbeito, G., Barbeito-Caamaño, C. e Bouzas-Mosquera, A. (2023b). A cardiotoxicity dataset for breast cancer patients. *Scientific Data*, 10, 527. https://doi.org/10.1038/s41597-023-02419-1.

Piñeiro-Lamas, B., López-Cheda, A. e Cao, R. (2025) `sicure`: *Single-Index Mixture Cure Models*. R package version 0.1.1. https://CRAN.R-project.org/package=sicure.

Piñeiro-Lamas, B., López-Cheda, A. e Cao, R. (2026a). Maximum likelihood estimation in single-index mixture cure models. *Submitted*.

Piñeiro-Lamas, B., López-Cheda, A. e Cao, R. (2026b). Single-index mixture cure models for functional covariates. *Submitted*.

Strzalkowska-Kominiak, E. e Cao, R. (2013). Maximum likelihood estimation for conditional distribution single-index models under censoring. *Journal of Multivariate Analysis*, 114, 74–98. https://doi.org/10.1016/j.jmva.2012.07.012.

Xu, J. e Peng, Y. (2014). Nonparametric cure rate estimation with covariates. *Canadian Journal of Statistics*, 42(1), 1–17. https://doi.org/10.1002/cjs.11197.

Sesión invitada da SGAPEIO no congreso SEIO 2026

por Marta Sestelo

A SGAPEIO estivo presente cunha sesión invitada no *XLII Congreso Nacional de Estadística, Investigación Operativa y Ciencia de Datos* (SEIO 2026), celebrado en Santiago de Compostela. A sesión celebrouse o mércores 2 de setembro de 2026, de 12:40 a 14:00 h, e foi moderada por Eva Soto-Antonio (socia da SGAPEIO e investigadora da UDC).

A sesión amosou a diversidade da investigación que se fai na comunidade galega de Estatística e Investigación Operativa. Os catro traballos presentados abranguen ámbitos moi diferentes: teoría de xogos, medidas de concordancia, estimación en áreas pequenas con datos funcionais e control estatístico de procesos. Os traballos presentados foron:

- **Un valor solidario desde una perspectiva global para juegos coalicionales**, de J. M. Alonso Meijide, M. Álvarez Mozos, M. G. Fiestras Janeiro e A. Jiménez Losada.
  Neste traballo propoñemos un valor para xogos globais (Gilboa e Lehrer, 1991) que verifica as propiedades de linealidade, anonimato e eficiencia, ao que chamamos *valor estrutural*. Asociando un xogo global a un xogo coalicional e utilizando o valor estrutural, definimos e caracterizamos un novo valor solidario para xogos coalicionais. Finalmente, comparamos este valor con outros valores solidarios (Nowak e Radzik, 1994; Casajus e Huettner, 2014).

- **Estimadores de diversos coeficientes delta en base al estimador insesgado de las proporciones esperadas de acuerdos**, de M. Álvarez Hernández e A. Martín Andrés.
  Para medir o grao de acordo entre dous observadores que clasifican independentemente a n suxeitos dentro de K categorías, adóitase utilizar o clásico coeficiente Kappa de Cohen. Dado que esta medida presenta defectos (como a súa mala actuación ante distribucións marxinais moi desequilibradas), en ocasións utilízase o coeficiente baseado no modelo de resposta Delta. Isto permite obter ademais a contribución de cada categoría ao acordo global e a súa consistencia. Recentemente comprobouse que o estimador clásico de Kappa de Cohen subestima o verdadeiro valor do parámetro e corrixiuse a devandita estimación. Neste traballo demóstrase que algo similar sucede na estimación do acordo global do modelo Delta e nos estimadores dos parámetros asociados. En consecuencia, propóñense novos estimadores menos nesgados, analízase o comportamento de todos eles e das súas varianzas, e conclúese que convén utilizar os novos estimadores cando n ou K son pequenos.

- **Estimation of fertility rates using functional data and small area estimation techniques**, de E. López Vizcaíno, M. J. Lombardía Cortiña e C. Rueda Sabater.
  Dispor de estimacións precisas da fecundidade municipal é fundamental para a planificación demográfica, pero as proxeccións en áreas pequenas son difíciles debido á escaseza e á alta variabilidade dos datos. Este estudo mellora a estimación das taxas de fecundidade dos concellos galegos entre 2002 e 2022 combinando a Análise de Datos Funcionais (FDA) e a Estimación en Áreas Pequenas (SAE). Propoñemos un novo modelo *Frequency Modulated Möbius* (FMM) que ofrece unha forma flexible e interpretable de captar as tendencias da fecundidade. O seu comportamento compárase co de métodos existentes mediante simulacións e datos reais. Os resultados mostran que os modelos baseados en FMM superan as alternativas, producindo curvas de fecundidade máis suaves e un menor RMSE. Nas predicións para 2023–2032, os modelos FMM tamén son os que mellor funcionan, e indican un descenso continuado da fecundidade e un atraso na idade á maternidade na maioría dos concellos. A validación cruzada confirma a robustez e a fiabilidade da proposta para a estimación e predición da fecundidade en áreas pequenas.

- **Monitorización del desempeño del proceso y detección de anomalías mediante gráficos de control de perfiles basados en residuos**, de E. Soto-Antonio, S. Naya, M. Francisco-Fernández e J. Tarrío-Saavedra.
  Amósase unha metodoloxía de control estatístico de procesos definidos por características críticas para a calidade dependentes de múltiples covariables, baseada en gráficos de control de perfís sobre residuos. Nunha fase de referencia axústase un modelo preditivo do perfil medio, considerando modelos lineais, GAM e métodos de aprendizaxe automática, e obtense a estrutura residual asocia-

da. Na fase de monitorización, os residuos analízanse mediante gráficos Shewhart e EWMA para detectar desviacións non explicadas pola relación previamente estimada entre variables. A metodoloxía ilústrase con datos de tempos de tránsito de buques Neo-Panamax na Canle de Panamá Ampliada e con datos reais de consumo enerxético en edificios. Os resultados amosan que este enfoque permite identificar cambios pequenos, patróns de aprendizaxe e cambios bruscos en sistemas complexos mediante un único gráfico de control, preservando a interpretación do efecto das covariables.

A sesión foi unha boa oportunidade para dar visibilidade ao traballo dos grupos de investigación vinculados á SGAPEIO no principal foro nacional da disciplina. Tamén serviu para seguir reforzando a colaboración entre ambas sociedades. Desde a SGAPEIO agradecemos a participación de todas as persoas relatoras e da moderadora, así como a acollida da organización do congreso.

Xubilación de Balbina Virginia Casas Méndez: Unha vida universitaria construída con compromiso e xenerosidade

por José Manuel Prada



A xubilación de Balbina Virginia Casas Méndez marca o final dunha longa etapa de dedicación á nosa Universidade e supón unha excelente oportunidade para recoñecer unha traxectoria profesional exemplar, construída día a día con traballo, responsabilidade e unha profunda vocación de servizo. Tras décadas de actividade académica, chega o momento de agradecer todo o que achegou á institución, aos seus estudantes e a quen tivemos a sorte de compartir con ela parte do noso camiño profesional.

Falar de Balbina é falar, en primeiro lugar, do seu compromiso coa docencia. Ao longo dos anos formou a numerosas xeracións de estudantes, transmitindo non só coñecementos, senón tamén valores esenciais para o exercicio profesional e para a vida. Quen asistiu ás súas clases lembra a súa capacidade para explicar con claridade, a súa proximidade e o seu interese por que cada estudante puidese desenvolver ao máximo as súas capacidades. Sempre entendeu o ensino como unha tarefa que ía máis aló das aulas, dedicando tempo e esforzo a orientar, aconsellar e acompañar a quen acudía a ela en busca de apoio.

A súa traxectoria académica estivo igualmente ligada a moitas outras responsabilidades universitarias. Como ocorre con tantas persoas que sosteñen o funcionamento diario das nosas institucións, Balbina participou activamente en tarefas de xestión, coordinación e organización que, aínda que a miúdo pasan desapercibidas, resultan fundamentais para o bo funcionamento da vida universitaria. En cada unha delas destacou polo seu sentido da responsabilidade, o seu rigor e a súa capacidade para afrontar os retos con serenidade e eficacia.

Pero se hai algo que quen a coñecemos valoramos especialmente é a súa disposición permanente para colaborar. Sempre estivo dispoñible para axudar cando xurdía unha dificultade, para compartir a súa experiencia cos compañeiros máis mozos ou para implicarse en calquera iniciativa que puidese contribuír a mellorar o noso contorno de traballo. A súa actitude positiva e construtiva converteu moitas situacións complexas en retos compartidos que se resolvían grazas ao esforzo colectivo.

Ao longo destes anos, Balbina foi tamén testemuña e protagonista dos numerosos cambios experimentados pola Universidade. Viu evolucionar titulacións, metodoloxías docentes, ferramentas tecnolóxicas e formas de traballo. En todos estes procesos demostrou unha notable capacidade de adaptación, afrontando as transformacións con intelixencia, espírito aberto e vontade de seguir aprendendo. Esa combinación de experiencia e capacidade de renovación constitúe, sen dúbida, un dos trazos máis destacados da súa traxectoria.

Con todo, reducir o seu legado unicamente ás súas contribucións profesionais sería quedarse curto. As institucións están formadas por persoas e son as relacións humanas as que terminan dando verdadeiro sentido ao traballo cotián. Neste ámbito, Balbina deixa unha pegada especialmente valiosa. A súa amabilidade, o seu sentido do compañeirismo, a súa xenerosidade e a súa capacidade para crear un ambiente cordial contribuíron a fortalecer os lazos entre todos nós. Para moitos foi unha compañeira de confianza; para outros, unha referencia profesional; para todos, unha persoa próxima coa que sempre resultou fácil traballar e compartir experiencias.

Por iso, aínda que a súa xubilación supón o peche dunha etapa importante para a nosa Universidade, tamén constitúe unha ocasión para celebrar unha carreira desenvolvida con dedicación e coherencia. Os proxectos nos que participou, os estudantes que formou e as persoas coas que colaborou son o mellor reflexo dunha traxectoria que deixou unha pegada duradeira.

Queremos expresarlle o noso máis sincero agradecemento por todos estes anos de traballo, compromiso e amizade. Estamos seguros de que esta nova etapa lle permitirá gozar con máis tempo das súas afeccións, da súa familia e daqueles proxectos persoais que quizais quedaron pospostos polas esixencias da vida académica.

A Universidade perde hoxe unha grande profesional en activo, pero conserva para sempre a contribución dunha compañeira que axudou a construíla e mellorala. Por todo iso, desexámoslle unha xubilación plena, feliz e merecida, coa certeza de que a lembranza do seu paso pola nosa institución permanecerá entre nós durante moitos anos.

Xubilación de Wenceslao González Manteiga

por Manuel Febrero Bande



O pasado 31 de agosto produciuse a xubilación de Wenceslao González Manteiga (Wences) como catedrático de Estatística e Investigación Operativa do departamento de Estatística, Análise Matemática e Optimización da Universidade de Santiago de Compostela e do seu nomeamento como profesor emérito. Atrás déixanse 47 anos adicados á Estatística na USC que vai ser difícil resumir nunhas poucas liñas.

O xeito máis rápido sería dicir que se xubilou (ou aposentou como se di no Brasil onde naceu) o mellor estatístico nacional de tódolos tempos (o GOAT ou "cabra" que se di agora) pero podería parecer que me deixo levar pola amizade de máis de 36 anos ou pola exaxeración propia destes escritos. Non hai exaxeración algunha. Como xa se deu conta neste mesmo boletín (ver Informest 68) a traxectoria investigadora de Wences, simplemente ao peso, non ten comparativa nacional nin en número de publicacións, nin en diversidade de temáticas, nin mesmo en proxectos aplicados ou competitivos nin no tamaño da rede de coautores. Pero non é hoxe o día de volver a contar as excelencias investigadoras de Wences que son tantas e tan diversas, que non teríamos espazo nin se lle faría boa xustiza senón, para lembrar todos estes anos adicados á institución da que se xubila.

O comezo desta andaina de 47 anos foi azaroso porque moi ben Wences estivo a piques de non adicarse á Estatística. Ocorreu que logo de licenciarse (tamén azarosa foi a licenciatura por outros motivos) foi chamado polo decano de Matemáticas para dar clases de Análise Matemática no campus de Lugo (–preséntese o vindeiro luns–, lle dixeron) pero, nun xiro de guión afortunado para as vindeiras xeracións de estatísticos galegos, o recentemente creado departamento de Estatística e Investigación Operativa ofreceulle unha praza en Santiago que el aceptou antepoñendo o benestar familiar a outras consideracións (o novo departamento era unha aposta e unha incerteza). Na cerna desta decisión está a motivación vital de Wences: a familia e non só na dimensión estritamente persoal. Construír practicamente da nada unha escola recoñecida internacionalmente e formar as 37 doutorandas/os que hoxe ocupan postos de responsabilidade na universidade (dous reitores galegos actualmente) ou nas empresas, require, non só paixón e alta capacidade de traballo senón argallar arredor de si unha comunidade de colaboradores cos mesmos valores de traballo, tesón, esforzo e humildade que aprendidos na familia propia estendeu á súa familia académica. Wences participou activamente no deseño e modernización das materias de Estatística dos diferentes plans de estudos de Matemáticas (licenciatura e grao) incorporando ideas ou temáticas que traía de congresos ou visitas internacionais. No seu historial docente en licenciatura e grao figuran sobre todo tópicos de Inferencia Estatística e Estatística Matemática aínda que elaborou materiais de moitas outras que xenerosamente compartía cos seus compañeiros. Se fixeches Matemáticas, con probabilidade 1/2 Wences foi o voso profesor na materia de 3º (compartía a docencia dos dous grupos con José Manuel Prada). Os anos adicados a ser profesor da UNED en Pontevedra, algúns anos con clases ateigadas de alumnos (non é o normal na UNED) de certas titulacións que buscaban entender e arranxar unha mala experiencia previa na docencia da Estatística na súa universidade de orixe, tamén axudou a que Wences soubera case que de todo. Nestes 47 anos tivo ofertas tentadoras para cambiar de aires e seguir a súa carreira noutros destinos onde, de seguro, tería máis recoñecemento pero, de novo quero pensar, que a motivación familiar, persoal e académica, impúxose sobre outras razóns. Por semellante motivo, non se prodigou moito na xestión universitaria fóra do ámbito estritamente académico. Foi director do departamento de Estatística e Investigación Operativa pero cando foi tentado para ocupar postos de xestión de máis alto rango, declinou amablemente a proposta vendo que iso o apartaba da súa principal paixón investigadora e docente. Egoistamente, coido que atinou.

O legado máis importante de Wences é a creación dunha canteira galega de estatísticos e facela recoñecida no mundo aplicando a súa inmensa capacidade para que se desenvolvan as mellores potencialidades dos seus alumnos ou colegas a base de liderar co propio exemplo e aplicando altas doses de paciencia e amabilidade (de novo, receitas dunha boa dinámica familiar). A receita parece simple pero non o é, require de apoio constante e xeneroso. Wences sempre está disposto a comentar e dar bos consellos sobre un traballo científico, a entusiasmarse por un recente resultado ou técnica que atopou, a celebrar os teus éxitos académicos e persoais coma se fosen propios ou mesmo a falar de calquera tema de actuali-

dade diante dun café e un *cake* (disto último que non se entere a súa filla Lidia). Gardo especialmente unha agarimosa lembranza das longas andainas, ás que Wences é tan afeccionado en sitios tan afastados da casa como Brasil, Perú, Sudáfrica ou Xapón. Nas conversas destas andainas soen aparecer as outras paixóns de Wences fóra da familia e a Estatística: Brasil e o fútbol ou viceversa, como recordatorio daquel rapaz, fillo de emigrantes galegos en São Paulo e que retornou a Galicia con 13 anos.

Grazas polas túas ensinanzas como profesor, mestre, guía e amigo e grazas polo teu exemplo constante que nos fai preguntarnos ante calquera dilema "Como o faría Wences?" e da resposta sempre aparece a boa solución. Foi o noso privilexio acompañarte nesta viaxe que aínda non rematou, porque o paso a emérito, como se amosa nestes meses, só fixo acrecentar o teu entusiasmo pola investigación e porque como se dixo no congreso deste pasado xullo onde celebramos a túa xubilación, "os investigadores non se xubilan, simplemente libéranse".



Wences e eu en Ipanema, Rio de Janeiro.

Medalla de Investigación da Real Academia Galega de Ciencias a Rosa M. Crujeiras Casais

por Pedro Faraldo



Rosa M. Crujeiras Casais (Artes-Ribeira 1978), catedrática de Estatística e Investigación Operativa na Universidade de Santiago de Compostela (USC), vén de recibir recentemente a distinción **Medalla de Investigación da Real Academia Galega de Ciencias (RAGC)** na categoría Enrique Vidal Abascal, da sección de Matemáticas, Física e Ciencias da Computación. Esta distinción *recoñece a traxectoria de científicos e científicas en activo, con actividade consolidada, desenvolvida maioritariamente en Galicia.* É o recoñecemento a unha vida dedicada a facer mellor a nosa ciencia, inspirando a novas xeracións e situando a nosa investigación na elite mundial. Sen dúbida, Rosa é máis que merecedora deste mérito, tanto polo seu gran talento como pola súa calidade como persoa.

Neste caso, recoñécese traxectoria investigadora con contribucións de referencia internacional no campo da Estatística e cuxo traballo deu lugar a importantes avances metodolóxicos, liderou proxectos competitivos de investigación, tanto no marco do Plan Nacional de I+D+i como impulsando unha notable actividade de transferencia de coñecemento mediante contratos con empresas e colaboracións con administracións públicas (en particular coa Xunta a través de estudos de inserción laboral e turismo). Así mesmo no desenvolvemento de ferramentas de software amplamente utilizadas pola comunidade investigadora. Tamén se recoñece o seu labor como directora científica do CITMAga desde 2021 contribuíndo de forma decisiva á súa creación e consolidación como Centro de referencia en Matemáticas. Tamén a RAGC valora o seu papel fundamental na difusión e divulgación da cultura científica, xunto á promoción da muller no eido das matemáticas.

Ao longo destes anos compartimos proxectos de investigación, materias, congresos, publicacións..., en definitiva, innumerables horas de traballo; pero tamén compartimos ilusións, preocupacións, celebracións e decepcións. En todos eses espazos, Rosa destacou sempre pola súa capacidade para comprender as persoas, construír equipos e facer que cada membro se sentise parte dun proxecto común. A súa forma de liderar sempre se baseou na confianza, na escoita e no respecto.

Os que a coñecemos ben sabemos a paixón que lle pon a cada proxecto ou iniciativa, xa sexa na docencia como na investigación, e podo dicir que tamén na xestión. É desas persoas que se comprometen activamente coa súa universidade, tendo asumidos diversos cargos de xestión como a dirección científica do Centro de Investigación e Tecnoloxía Matemática de Galicia (CITMAga), ata dar o paso á máxima responsabilidade no goberno académico, a propia Reitoría da Universidade de Santiago de Compostela.

No eido científico, a súa traxectoria é verdadeiramente frutuosa e referente, centrada no desenvolvemento de metodoloxías non paramétricas para a análise de datos en contextos complexos, concretamente en procesos que implican dependencia espacial e variables observadas en espazos non euclidianos (como datos circulares e direccionais). Todo o seu rigor levouna a publicar en revistas estatísticas de gran prestixio incluíndo *Annals of Applied Statistics* (2023), *The Journal of the American Statistical Association* (2024) ou *Biometrics* (2023), entre outras. Con estes traballos, acadou resultados salientables no desenvolvemento de novos métodos para a análise de datos espaciais, de datos circulares e direccionais, con contribucións en ámbitos como a análise espectral de procesos espaciais, a estimación non paramétrica, a regresión e a análise da multimodalidade. Ademais, dirixiu sete teses de doutoramento e exerceu como investigadora principal (ou co-investigadora principal) en tres proxectos nacionais. A todo isto súmase a súa intensa actividade de transferencia de coñecemento, coa participación en máis de corenta iniciativas con empresas e entidades públicas, xa sexa como investigadora principal ou como membro do equipo do proxecto.

A súa traxectoria ten un marcado carácter internacional e interdisciplinar. Colaborou con grupos de investigación de Bélxica, Reino Unido, Italia e Portugal, e desenvolveu aplicacións de metodoloxías estatísticas en eidos tan variados como a ecoloxía, a enxeñaría química, a lingüística, a economía ou a informática. Esta investigación transversal tradúcese no desenvolvemento de software libre, sendo coautora de paquetes de R como *DTDA*, *multimode*, *NPCirc* e *HDiR*, ferramentas que permiten

aplicar a estatística avanzada a problemas reais da nosa sociedade. A súa calidade investigadora reflíctese tamén na súa presenza nos consellos editoriais de revistas internacionais de referencia, coma o *Journal of Nonparametric Statistics, Environmetrics, Spatial Statistics* ou o *Journal of the Royal Statistical Society: Series C*.

Ao longo da súa carreira, diversas distincións recoñeceron a gran proxección da profesora Rosa Crujeiras. En 2023 recibiu o Premio ao *Talento Científico Feminino Novo* da Fundación da Real Academia de Ciencias Exactas, Físicas e Naturais. No ano 2024, xunto coas profesoras M. Alonso e I. Gijbels, obtivo o Premio da Sociedad de Estadística e Investigación Operativa (SEIO) - Fundación BBVA á mellor contribución metodolóxica en estatística.

Rosa, sentimos unha enorme satisfacción e un profundo orgullo ao ver recoñecida a túa traxectoria, o teu compromiso por facer ciencia de excelencia e a túa extraordinaria capacidade de comunicar a relevancia dos resultados, tanto á comunidade científica como ao conxunto da sociedade.

O seu mestre da Escola Rural Agrupada de Artes, xa dicía de Rosa aos 12 anos que era unha alumna excelente en todos os aspectos e que posuía unha capacidade intelectual extraordinaria para a súa idade.

Estamos totalmente seguros de que seguirás abordando novos retos e cruzando novas fronteiras.

Moitísimas felicidades!



VI PhDay-EIO e premio da SGAPEIO á mellor comunicación oral

por Laura Davila Pena

O pasado 23 de outubro de 2026 celebrouse na Facultade de Ciencias Económicas e Empresariais da Universidade de Vigo (UVigo) o VI PhDay-EIO, unha xornada dirixida ao alumnado do Programa de Doutoramento en Estatística e Investigación Operativa (PDEIO) das tres universidades públicas galegas.

A xornada comezou cun acto de apertura no que participaron Jacobo de Uña Álvarez (Grupo SiDOR), Leticia Lorenzo Picado (Máster en Técnicas Estatísticas), María Xosé Rodríguez Álvarez (Departamento de Estatística e Investigación Operativa da UVigo) e Juan Vidal Puga (PDEIO e SGAPEIO), en representación das distintas entidades implicadas. No evento houbo sete relatorios científicos de alta calidade a cargo de doutorandas e doutorandos da Universidade da Coruña (UDC), da Universidade de Santiago de Compostela (USC) e da UVigo. Abordaron temas tan variados como a monitorización de datos funcionais, a estimación de hipersuperficies, a análise de supervivencia, a optimización enteira mixta, os datos faltantes, a aprendizaxe por reforzo en sistemas biolóxicos ou a teoría de xogos.

Un ano máis, a xornada contou co premio da SGAPEIO á mellor comunicación oral. Nesta ocasión, recibiuno Héctor González Vázquez (USC) pola súa presentación titulada "De datos con ruído a variedades: estimación minimax de hipersuperficies". O xurado estivo formado por María Xosé Rodríguez Álvarez (presidenta), Laura Freijeiro González (secretaria) e Brais González Rodríguez (vogal), e o premio foi entregado por Nora M. Villanueva, socia da SGAPEIO. Dende aquí: moitos parabéns, Héctor!



De esquerda a dereita: Nora M. Villanueva (socia da SGAPEIO e encargada da entrega do premio), Paula Soto Rodríguez (membro do Comité Organizador do VI PhDay-EIO), Héctor González Vázquez (gañador do premio) e Alejandro Bernárdez Ferradás (membro do Comité Organizador do VI PhDay-EIO).

O acto contou ademais coa sesión "Tendendo pontes entre universidade e empresa", moderada por Brais González Rodríguez (UVigo) e na que participaron Nora M. Villanueva (UVigo) e Ángel Manuel González Rueda (USC). Quen asistiu tivo tamén a oportunidade de coñecer de primeira man distintas traxectorias despois do doutoramento, tanto dentro como fóra da academia, na mesa redonda "A carreira investigadora, sen filtros", moderada por Alejandro Bernárdez Ferradás (UVigo) e na que participaron Adrián Lago Balseiro (UVigo), Laura Freijeiro González (UVigo), Nora M. Villanueva (UVigo) e Adriana Navarro Ramos (Universidade de Granada).

O VI PhDay-EIO foi organizado por Alejandro Bernárdez Ferradás e Paula Soto Rodríguez, doutorandos do PDEIO na UVigo, e contou co apoio do PDEIO, do Grupo de Investigación SiDOR, do Departamento de Estatística e Investigación Operativa da UVigo, do Máster en Técnicas Estatísticas e da SGAPEIO. A xornada súmase ao éxito das edicións anteriores e, polo que contan as persoas asistentes, Paula e Álex puxeron o listón moi alto. Parabéns!



Fotografía de grupo coas persoas asistentes ao VI PhDay-EIO.

Novidades na información estatística publicada

Conta satélite do deporte

A Conta Satélite do Deporte é unha actividade estatística de síntese destinada a medir a dimensión económica do sector do deporte en Galicia e a súa contribución á economía galega. Este proxecto xurdiu a iniciativa da Secretaría Xeral do Deporte (Consellería de Presidencia, Xustiza e Deportes), que solicitou ao IGE a cuantificación da dimensión económica do sector.

A definición estatística do sector do deporte non está regulada en Europa, aínda que se toma como referencia a Clasificación de Vilnius, baseada en estándares estatísticos.

Esta publicación ofrece resultados da conta de produción, explotación e emprego do sector do deporte correspondentes ao período 2021-2024, en coherencia co Sistema de Contas Económicas de Galicia (Revisión estatística 2024).

No ano 2024, o valor engadido bruto xerado polo sector do deporte en Galicia ascendeu a 731.245 miles de euros, o que representa o 0,90 % do Produto Interior Bruto (PIB) xerado na Comunidade autónoma en dito período. Con respecto a 2023, o valor xerado a prezos correntes polo sector medrou a unha taxa do 2 %.

En termos de emprego, o sector é responsable directo de 17.205 postos de traballo, que representan o 1,50 % do total do emprego da economía galega. + info

Enquisa estrutural a fogares. Mobilidade

A Enquisa estrutural a fogares recolle cada ano información sobre os fogares galegos e as súas características socioeconómicas. Consta de dúas partes: unha primeira de carácter xeral que se repite todos os anos, deseñada para recoller información sobre variables básicas, e unha segunda de carácter específico. No ano 2025 a parte específica dedicouse á Mobilidade. O principal obxectivo deste módulo é obter información sobre as pautas de mobilidade da poboación galega.

A mobilidade por traballo en Galicia caracterízase polo predominio do transporte privado, a proximidade entre a vivenda e o centro de traballo e unha presenza aínda limitada do teletraballo. Máis da metade das persoas ocupadas traballa no mesmo concello no que reside, case tres de cada catro das que se desprazan percorren menos de 15 km e o 10,83 % teletraballou polo menos unha vez á semana en 2025.

Por outra parte, a maioría das persoas que estudan en Galicia desprázanse andando, en bicicleta ou en transporte público ao seu centro de estudos. Na maioría dos casos, o centro de estudos atópase no mesmo concello que o da súa residencia, chegando en menos de 15 minutos ao seu destino.

Para os desprazamentos distintos do traballo e do estudo o transporte privado é o medio máis usado. Inclúense aquí desprazamentos para compras, ocio, visitas a familiares e amigos, acompañamento de nenos ou maiores e xestións persoais ou familiares. A porcentaxe de uso do transporte privado rolda entre o 56,47 % para ir de ocio e o 67,02 % para facer visitas a familiares e amigos. + info

Escenario macroeconómico

O IGE difunde o escenario macroeconómico para Galicia de horizonte quinquenal. Este escenario refírese á evolución prevista do PIB e dos seus compoñentes desde a óptica da demanda: consumo privado, consumo público, formación bruta de capital, exportacións e importacións de bens e servizos. O cadro complétase coas previsións do emprego: postos de traballo equivalentes a tempo completo e taxa de paro.

A información de base para elaborar as previsións son as Contas económicas anuais e as Contas económicas trimestrais elaboradas polo IGE.

Segundo o escenario macroeconómico, o PIB galego crecerá un 1,9 % no ano 2027 e a taxa de paro en Galicia reducirase ata o 7,7 % en 2027. + info

Novas institucionais

- Nos meses de xuño e outubro celebráronse dúas reunións do Consello Reitor do IGE. Na primeira aprobáronse a conta de liquidación do IGE correspondente ao exercicio 2025 e o informe do Plan anual de actuación dese mesmo ano, mentres que na segunda se aprobou o anteproxecto de orzamento para o exercicio 2027.

- En xullo asinouse un acordo de colaboración coa Consellería de Medio Ambiente e Cambio Climático e a Consellería de Medio Rural para o estudo estatístico do desperdicio alimentario en Galicia.

- No mesmo mes, o día 30, celebrouse unha reunión do Pleno do Consello Galego de Estatística, na que se informaron favorablemente o anteproxecto de Lei do Plan Galego de Estatística 2027-2031 e o proxecto do novo Estatuto do IGE. Ambos os documentos atópanse actualmente en fase de exposición pública. Así mesmo, durante a reunión presentouse unha versión beta da nova páxina web do IGE.

Traballos defendidos no SUG

Traballos Fin de Mestrado

Análise de datos agregados por área: contaminación atmosférica e desigualdade socioeconómica

Diana Cristina Castro Armenta
Directora e director: María Xosé Rodríguez Álvarez e Tomás Raimundo Cotos Yáñez
USC. Modalidade académica

Segundo as análises da Axencia Europea de Medio Ambiente (AEMA), en 2022 a exposición a NO2 e PM2.5 foi responsable dun elevado número de mortes prematuras en Europa. Desde unha perspectiva de xustiza ambiental, é relevante estudar se a exposición á contaminación atmosférica está distribuída de forma desigual entre a poboación, especialmente en relación co nivel socioeconómico das zonas residenciais. Este estudo aplica ferramentas de análise de datos agregados por área para examinar esta cuestión no contexto da comunidade autónoma de Galicia. En concreto, analiza a posible asociación entre os niveis de NO2 e PM2.5 e a renda per cápita municipal, co obxectivo de avaliar se existen patróns territoriais consistentes cunha distribución desigual da exposición a estes contaminantes segundo o nivel de renda. Para iso, utilízanse técnicas descritivas, cartográficas e estatísticas axeitadas para datos agregados por área, tendo en conta a estrutura espacial da información dispoñible.

Modelos econométricos e risco de crédito: Desenvolvemento dunha aplicación web en Shiny e proposta de enfoques alternativos

María Iglesias Martínez
Directoras: Marta Sestelo Pérez e Nora Martínez Villanueva
Titores: Fernando Álvarez e Alberto González Almuíña (ABANCA)
UVigo. Modalidade empresa

Este Traballo Fin de Máster desenvolve unha metodoloxía econométrica para estimar a probabilidade de incumprimento ao longo do ciclo económico nunha carteira hipotecaria da entidade financeira ABANCA. O traballo enmárcase no proxecto ALICE e utiliza modelos lineais con variables macroeconómicas, financeiras e internas para realizar unha proxección retrospectiva da frecuencia observada de incumprimento agregada. A metodoloxía combina enxeñaría de variables, filtros econométricos, validación estatística, coherencia económica e un ranking multicriterio para seleccionar modelos interpretables e trazables. Os resultados mostran que, a partir dun universo amplo de variables e combinacións, é posible obter un conxunto reducido de modelos economicamente coherentes. Ademais, desenvólvese unha aplicación web en Shiny, baseada na linguaxe de programación R, para facilitar a visualización e revisión dos resultados. Finalmente, compáranse os modelos principais con enfoques alternativos baseados en regresión penalizada, que actúan como contraste complementario da metodoloxía proposta.

Estudo, análise e implementación de técnicas de detección de *data drift* e *concept drift* para ciberseguridade

Marcos Montaña Fernández
Directora: Marta Sestelo Pérez
Titor: Pablo Menéndez Trillo (GRADIANT)
USC. Modalidade empresa

Este traballo estuda e implementa técnicas de detección de *data drift* e *concept drift* aplicadas á ciberseguridade, concretamente á detección de anomalías. O obxectivo é avaliar e comparar de forma sistemática distintos métodos de detección, ante a necesidade de anticipar a degradación dos modelos de aprendizaxe automática cando os datos ou os patróns aprendidos deixan de representar a realidade. Para iso deseñouse unha metodoloxía experimental baseada na inxección sintética de distintos tipos e intensidades de *drift*, permitindo analizar o comportamento dos métodos en escenarios controlados. O *data drift* abórdase dende unha perspectiva univariante e multivariante, mentres que o *concept drift*, dada a dificultade da súa observación directa, analízase mediante sinais derivados dun modelo de detección de anomalías baseado nun Autoencoder. Os resultados permiten identificar as fortalezas e limitacións de cada técnica e ofrecen criterios prácticos para seleccionar a máis axeitada segundo o escenario considerado.

Modelización da actividade económica española en tempo real: un enfoque de *Nowcasting* mediante Modelos de Factores Dinámicos

Yago Núñez García
Directores: Rubén Fernández Casal e Germán Aneiros Pérez
Titoras e titor: Teresa Veiga Rodríguez, Ana Blanco Bugueiro e Sergio Díaz Canosa (ABANCA)
UDC. Modalidade empresa

Anticipar os puntos de xiro do ciclo económico é clave para a toma de decisións no sector bancario, pero o Produto Interior Bruto, principal indicador da actividade española, publícase cun desfasamento de trinta días. Para dar resposta, deseñouse un indicador de *nowcasting* baseado nun Modelo Factorial Dinámico no espazo de estados, estimado mediante o algoritmo EM e Filtro de Kalman, que extrae un factor común dun panel de indicadores de frecuencia mixta. Un pipeline automatizado para preprocesar as series (desestacionalización, estacionariedade e estandarización) e un tratamento do período atípico da COVID-19. As estimacións valídanse fóra de mostra mediante xanela rodante, mostrando unha mellora preditiva fronte á inercia histórica (AR(1)) e previsións doutros organismos. Intégrase nunha aplicación web en R Shiny cun asistente virtual baseado nunha arquitectura RAG, que fragmenta e procesa a memoria do proxecto (*chunking*) para guiar o usuario e resolver dúbidas, democratizando o seu uso sen coñecementos técnicos.

Aplicación de técnicas estatísticas no estudo da relación entre riqueza e temperatura nos EE.UU.

Jorge Pazó Zurrón
Directora: María Isabel Borrajo García
Titora: María Isabel Neira Gómez (ECOBAS)
USC. Modalidade empresa

Este traballo analiza a relación entre temperatura e pobreza en distritos censais urbanos dos Estados Unidos mediante técnicas de estimación en áreas pequenas (SAE). O estudo integra información socioeconómica da American Community Survey con datos sobre a temperatura superficial e a do aire, construíndo unha base de datos de 53.920 distritos censais.

Como metodoloxía principal empréganse modelos Fay–Herriot con transformación loxística, obtendo estimacións EBLUP que melloran a precisión das estimacións directas de pobreza. Ademais, incorpóranse variables demográficas, educativas, residenciais, de mobilidade, desigualdade e orixe étnica-racial para explicar as diferenzas territoriais.

Os resultados mostran unha mellora na calidade das estimacións e evidencian unha asociación, aínda que reducida, positiva entre temperaturas máis elevadas e maiores niveis de pobreza. Así mesmo, a comparación entre temperatura do aire e temperatura superficial revela diferenzas mínimas, o que suxire que ambos os indicadores ofrecen un comportamento moi similar neste contexto urbano.

Cooperación no problema do carteiro chinés

María Gabriela Pérez Sierra
Directora: Silvia María Lorenzo Freire
UDC. Modalidade académica

Este traballo aborda o Problema do Carteiro Chinés (CPP) integrando a optimización combinatoria e a Teoría de Xogos Cooperativos para resolver o deseño de rutas de mantemento e a asignación xusta de custos entre axentes. Desenvolveuse a metodoloxía clásica de eulerización mediante o algoritmo de *Blossom* (Edmonds, 1965) e a regra de reparto γ (Hamers et al., 1999) en grafos debilmente eulerianos, garantindo solucións pertencentes ao *Core*. Ademais, desenvolveuse o paquete nativo de R CoopPostman, que resolve o emparellamento óptimo e o cálculo do xogo cooperativo de forma autónoma. A metodoloxía aplicouse con éxito a dúas redes reais da Coruña (Rolda de Outeiro e Obelisco–María Pita con subsidio municipal), validando a súa equidade e eficiencia computacional.

Implementación en R de regras de reparto na árbore de custo mínimo

Lucía Souto Vázquez
Directores: Gustavo Bergantiños Cid e Alejandro Saavedra Nieves
USC. Modalidade académica

O problema da árbore de custo mínimo (MCSTP, polas súas siglas en inglés) é un dos problemas clásicos da investigación operativa. Consiste en conectar un conxunto de axentes a unha fonte común minimizando o custo total da rede. Nos últimos anos, abordouse un novo aspecto: como repartir ese custo entre os diversos axentes. Para iso, propuxéronse varias regras de reparto que se poden dividir en dúas grandes familias: as regras baseadas nos algoritmos clásicos de construción da árbore e as regras baseadas nos xogos cooperativos. Neste traballo revísanse as principais regras de reparto existentes e preséntase 'mcstprules', un paquete de R que implementa, compara e analiza estas regras, incorporando ferramentas para estudar a estabilidade dos repartos. Finalmente, avalíase o rendemento computacional do paquete, comparando o cálculo exacto de certas regras cunha aproximación baseada no método de Monte Carlo, útil para problemas de gran tamaño onde o cálculo exacto é inviable.

Enfoque cooperativo do xogo de asignación

Johanna Karewith Vazquez López
Directora: Leticia Lorenzo Picado
UDC. Modalidade académica

Este traballo estuda o problema de asignación desde un enfoque cooperativo. En primeiro lugar, preséntase como un modelo de optimización mediante programación lineal, analizando algúns métodos de resolución e o papel do problema dual, cuxas variables poden interpretarse como pagamentos ou prezos asociados aos axentes. A partir desta base, o problema analízase como un xogo cooperativo con utilidade transferible. En particular, estúdase o xogo de asignación, no que compradores e vendedores poden formar coalicións e xerar valor mediante emparellamentos. O núcleo ocupa un lugar central, ao recoller os repartos estables do beneficio total. Ademais, considéranse outros conceptos de solución, como o valor de Shapley, o τ-valor e o nucleolo. Finalmente, desenvólvese un procedemento en R para construír un xogo de asignación a partir dunha matriz de beneficios e calcular algunhas das súas solucións cooperativas.

Traballos Fin de Grao

Análise estatística en datos con truncamento

Miguel Diz Castro
Titora: Mercedes Conde Amboage
Titulación: Grao en Matemáticas. USC

Este Traballo Fin de Grao aborda metodoloxías de estimación estatística non paramétrica para variables aleatorias afectadas polo fenómeno do truncamento aleatorio pola esquerda, habitual no contexto da Análise de Supervivencia, cuxas observacións situadas por debaixo dun certo limiar non chegan a rexistrarse.

Mediante estudos de simulación, demóstrase que os estimadores tradicionais, como a función de distribución empírica e o estimador tipo núcleo da función de densidade, son inconsistentes cando se aplican a datos truncados, xerando nesgos estatísticos severos.

Para superar estas limitacións, estúdase o estimador de Lynden-Bell para a función de distribución, xunto co estimador tipo núcleo da función densidade para datos truncados mediante a incorporación dos pesos de Lynden-Bell. Analízanse as súas propiedades asintóticas, as condicións de consistencia e os métodos de selección do parámetro de suavizado para a estimación da función de densidade.

Finalmente, a metodoloxía introducida aplícase á base de datos abortion do paquete mvna do software de libre distribución R para analizar os posibles efectos da exposición á cumarina en mulleres embarazadas.

Modelo ANCOVA non paramétrico

Alejandra Iglesias Molares
Titora: María Alonso Pena
Titulación: Grao en Matemáticas. USC

Este traballo presenta o modelo de análise da covarianza (ANCOVA) non paramétrico como ferramenta para comparar as funcións de regresión de distintos grupos dunha variable categórica. En primeiro lugar, faise unha introdución á regresión linear, incluíndo os modelos de análise da varianza (ANOVA) e de análise da covarianza (ANCOVA). En segundo lugar, preséntase unha alternativa á regresión linear: a regresión polinómica local. A continuación, estúdase o modelo ANCOVA non paramétrico, analizando de xeito teórico os contrastes de igualdade e paralelismo das funcións de regresión de distintos grupos dunha variable categórica. Por último, co obxectivo de mostrar a utilidade que poden ter na vida real, aplícanse os modelos vistos ao longo do traballo a datos sobre ingresos mensuais e gastos anuais nos fogares de Ceuta e Melilla procedentes da Enquisa de Presupostos Familiares do Instituto Nacional de Estadística de 2024.

Aplicación de modelos multiestado para a análise da supervivencia en pacientes con leucemia mieloide aguda

Alba Míguez Fernández
Titor: Juan Carlos Pardo Fernández
Titulación: Grao en Enxeñaría Biomédica. UVigo

Este traballo enfócase no estudo da leucemia mieloide aguda (LMA), un cancro moi agresivo e de rápida evolución. O obxectivo principal é analizar a supervivencia global e como inflúen nesta os eventos intermedios tales como a recaída e outras variables como o tipo de tratamento ou mutación.

Para iso, empregarase a base de datos Myeloid proporcionada polo paquete survival de R, que simula unha mostra de 646 pacientes diagnosticados con LMA con distintas mutacións no xen FLT3. O estudo comeza dende técnicas máis convencionais para estimar a supervivencia global como o estimador non paramétrico de Kaplan-Meier, que ademais permite facer distincións de supervivencia entre grupos con distinto tipo de tratamento, sexo, etc. Dada a complexidade da enfermidade hai que ter en conta a relevancia dos eventos intermedios (como a recaída ou transplante) dende que o paciente entra ata que sae deste (xa sexa por morte ou censura). Desta forma, implementarase un modelo enfermidade-morte de tres estados sen recuperación, e aplicaranse probas de modelos multiestado para calcular probabilidades de transición.

Así conséguese demostrar que a análise de eventos intermedios é fundamental para un estudo completo e realista da supervivencia, ademais de poder observar o efecto doutras variables como o tipo de tratamento, o que nun contexto clínico resulta moi relevante.

Estudo do efecto de factores prognósticos na supervivencia de pacientes de carcinoma hepatocelular mediante modelos de riscos proporcionais

Francisco Rodríguez Bianchi
Titor: Juan Carlos Pardo Fernández
Titulación: Grao en Enxeñaría Biomédica. UVigo

Este traballo recolle detalladamente todos os pasos seguidos nun estudo da evolución de pacientes oncolóxicos. As técnicas utilizadas pertencen ao campo da análise de supervivencia e permiten atopar relacións entre características dos pacientes e o tempo transcorrido ata determinado evento, como pode ser a morte.

Especificamente, estúdase unha cohorte concreta recollida na base de datos do Cancer Genome Atlas do National Cancer Institute dos Estados Unidos. A base de datos está conformada por pacientes de carcinoma hepatocelular (HCC) que pasaron por unha cirurxía de resección. Este documento inclúe unha introdución aos aspectos fundamentais desta enfermidade, unha das principais causas de morte por cancro no mundo, e tamén información de interese sobre os procedementos de resección. Cada individuo vese caracterizado por case 100 variables. Antes do comezo da análise, defínese cales delas contan coas cualidades e/ou a relevancia científica necesarias para seren utilizadas como factores prognósticos.

Todas as técnicas matemáticas empregadas quedan explicadas neste documento, xunto cos conceptos estatísticos (censura, risco, etc.) necesarios para comprendelas. O estimador de Kaplan-Meier é útil para comprender o perfil de supervivencia global da cohorte. Pola outra parte, o modelo semiparamétrico de riscos proporcionais de Cox permite analizar como varían os resultados ao modificar os valores dunha variable e, aínda máis importante, contrastar os efectos producidos por dúas ou máis variables. A aplicación destas técnicas, así como o previo tratamento dos datos, lévase a cabo utilizando a linguaxe R. A análise dos resultados desde unha perspectiva clínica permite establecer as conclusións que se presentan no peche deste documento.

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 unha socia: Erika Francesch Domenech

Erika Francesch Domenech é investigadora vinculada ao CITIC e técnica de investigación na Área de Control de Calidade de Corporación Hijos de Rivera, onde desenvolve un doutoramento industrial en colaboración co grupo MODES e o Departamento de Matemáticas da Universidade da Coruña.



Descríbenos o teu traballo

Actualmente traballo como técnica de investigación na Área de Control de Calidade de Corporación Hijos de Rivera, onde desenvolvo un doutoramento industrial en colaboración coa Universidade da Coruña e financiado pola Axencia Estatal de Investigación. No meu día a día combino análise e modelización estatística, tanto no marco do doutoramento como dando apoio a outros proxectos, conectando investigación e aplicación práctica.

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

Levo traballando en investigación dende 2012. Comecei nun ámbito máis experimental e, pouco a pouco, a análise de datos e a estatística foron gañando protagonismo na miña traxectoria profesional.

Dende cando es socia da SGAPEIO?

Dende 2025. Son unha incorporación moi recente.

Cal foi o teu 1º Congreso da SGAPEIO?

Aínda non tiven a oportunidade de asistir a ningún, así que o seguinte será o primeiro!

O tema do teu próximo proxecto ou artigo?

Agora mesmo estou traballando en varios proxectos relacionados coa mellora de procesos na industria cervexeira, abordados dende unha perspectiva estatística.

Un curso que che gustaría facer?

Un curso sobre *Agentic AI* e as súas aplicacións.

Que libro de estatística ou I.O. recomendarías?

Applied Predictive Modeling, de Max Kuhn e Kjell Johnson. Paréceme unha referencia moi útil para achegar a modelización estatística a problemas reais.

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

Posiblemente seguiría vinculada á investigación, máis orientada á xestión e coordinación de proxectos multidisciplinares.

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

Poñer orde nos datos e facerlles boas preguntas.

Que é o que máis valoras do labor da SGAPEIO?

O que máis me gusta é a comunidade que se crea arredor da estatística e a oportunidade de coñecer persoas coas que compartir experiencias e diferentes formas de traballar. Tamén me parece especialmente interesante AVATAR, como iniciativa para achegar a estatística aos máis novos dunha forma atractiva.

En que cres que deberiamos mellorar?

Creo que se está facendo un traballo moi bo e con moita ilusión. Quizais, como oportunidade para seguir medrando, reforzaría aínda máis a conexión co ámbito empresarial e industrial e daría maior visibilidade ás aplicacións da estatística nestes contextos.

As partículas subatómicas non entenden de probabilidades

por Julián Costa Bouzas

É moi coñecida a anécdota do matemático francés d'Alembert, que en 1754 publicou en L'Encyclopédie que, se lanzas dúas moedas ao aire, os resultados posibles son: dúas caras, dúas cruces, e unha cara e unha cruz; e as probabilidades son $1/3$ en todos os casos. Pero, sabías que Satyendra Nathan Bose, un físico da India, cometeu o mesmo erro a principios da década de 1920, con resultados sorprendentes? Bose estaba a estudar a disposición de fotóns (partículas cuánticas) en estados de enerxía, un problema similar a repartir n bólas en r urnas (amplamente coñecido como o problema do aniversario). Por exemplo, se $n = 3$ e $r = 3$, hai 27 disposicións posibles (variacións con repetición), e os físicos crían naquel entón que as partículas se repartían sempre ao azar, dando lugar a que as 27 disposicións sexan equiprobables, propiedade coñecida como estatística (distribución) de Maxwell–Boltzmann na física. Pero como as partículas son indistinguibles, Bose pensou que só hai 10 disposicións posibles (combinacións con repetición), e non 27, e que todas teñen a probabilidade de $1/10$ (o mesmo que pensara d'Alembert dous séculos antes). A sorpresa, imaxino que maiúscula, foi que esta predición foi confirmada polos experimentos que realizou, e así se deu conta de que, á fin e ao cabo, o seu erro non fora tal, senón un afortunado acerto. Bose escribiu un artigo cos seus resultados, que enviou a Albert Einstein, dando lugar a unha frutífera colaboración coñecida como estatística de Bose–Einstein. Por primeira vez, comprobárase que a estatística de Maxwell–Boltzmann non era adecuada para todas as partículas microscópicas en todas as circunstancias. As partículas que seguen a estatística de Bose–Einstein son chamadas bosóns en honra a Bose, e son moi importantes na física subatómica. Finalmente, a teoría foi demostrada en laboratorio en 1995 por Cornell, Wieman e Ketterle, sendo merecedores do premio Nobel de física.

informest é unha publicación cuadrimestral da SGAPEIO. Podes atopar os números anteriores xa publicados en www.sgapeio.es/, no apartado PUBLICACIÓNS.

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:
Laura Davila Pena e Marta Sestelo Pérez.
Paula Blanco Mosquera, Mª Esther Calvo Ocampo, Mª Erika Francesch Domenech, Fabio José González Eiras, Ana López Cheda, Leticia Lorenzo Picado, Manuel Oviedo de la Fuente e Juan José Vidal Puga.

Revisión lingüística:
Trinidad Pazos Celis.

Coa colaboración de:
Héctor González Vázquez, Beatriz Piñeiro Lamas, Ana López Cheda, Ricardo Cao Abad, José Manuel Prada, Manuel Febrero Bande, Pedro Faraldo Roca, Esther López Vizcaíno e Julián Costa Bouzas.

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

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