

Colaboración

Software

Actividades da SGAPEIO

Novas do IGE

Traballos de Estatística e IO

Coñecéndonos

Estatística recreativa

## Editorial

JAVIER TARRÍO SAAVEDRA – VICEPRESIDENTE DE ESTATÍSTICA.

Así como adoita caer o sol no mar de Vigo, grande e intenso, vén de rematar o XIV Congreso da SGAPEIO, deixando en todos nós unha cálida sensación de camaradería, coñecemento e vivencias compartidas. Por tanto, antes de nada, neste luscofusco e desde a pequena xanela que temos no *Informest*, quería agradecer todo o inxente e eficiente traballo feito polo Comité Organizador e polo Comité Científico, desde a planificación das sesións e dos cursos á organización dos premios, pasando polo magnífico programa social. Fago extensivo o agradecemento a todos os participantes, por marcar nas vosas axendas estas datas, por facer posible un congreso de alta calidade, que, partindo do ámbito autonómico ten vocación nacional e internacional.

Co agradecemento ao grupo Sidor nos labios, cruzo o Berbés, baixo ao Areal, subo a Urzaiz e collo o tren de volta ao norte coa impresión de termos creado e, o que é máis difícil, mantido, un proxecto único, unha familia cuxo vínculo é unha mesma formación e un xeito similar de entender a Estatística e a súa aplicación. E medra a certidume de que nesta horta vén agromando unha serie de froitos entre os que destacan a nosa propia sociedade, a SGAPEIO, titulacións comúns ás tres universidades galegas como o Mestrado en Técnicas Estatísticas e un programa de doutoramento conxunto, entre outros moitos fitos como, por exemplo, o ITMATI, que transcenden a conxuntura de competencia entre universidades para aportar, repito, un substrato, rede, tecido único en España, conectado como poucos co resto do mundo.

Dou en pensar que o noso proxecto-equipo-xeito de ver a Estatística é un balandro como os que vemos na Estación Marítima de Vigo perfeccionado durante anos para navegar por un determinado mar, cunhas condicións específicas. . . Non obstante, dun tempo a esta parte, o océano polo que navegamos está a mudar, cambian os ventos e as correntes, logo tamén crece a incertidume de chegar a porto. Este é o novo marco para a profesión que nos presentou Daniel Peña, no Congreso da SEIO en Alcoi. Foi a necesidade de resolver problemas prácticos o que fixo da Estatística, da man da ciencia de datos, unha ferramenta necesaria para a maioría dos perfís profesionais da actualidade. E por isto mesmo, non provocando nós o cambio, é preciso que nos adaptemos a el. Dicía Daniel Peña que é fundamental abrirse a resolver, desde o rigor do paradigma estatístico, aqueles problemas que están afrontando desde outras ramas da ciencia, fundamentalmente, desde as ciencias da computación. A modelización dos novos datos, sexan son, vídeos, imaxes, a combinación de todos, é fundamental para o avance da Estatística tal e como a coñecemos, e só se poderá facer saíndo da nosa zona de confort, incluíndo o eido da docencia. De feito, os novos profesionais da Estatística deberían incorporar capacidades computacionais (*machine learning*, bases de datos), pois, se ben hai anos os estatísticos traballaban como informáticos, son os informáticos de hoxe os que están traballando como estatísticos. Polo tanto, cómpre unha transformación, atinguindo incluso aos cursos básicos de Estatística en titulacións aplicadas, que deberíamos orientar a resolver os problemas específicos a través ferramentas multivariantes, clúster, dependencia. . . baixando á realidade de cada rama do saber, aínda que sexa sacrificando, en certa medida, o rigor formal ao que estamos afeitos. En Galicia creo que tomamos este camiño, sirvan de mostra as novas titulacións relacionadas coa ciencia de datos, os cambios nos plans de estudos. . . E estou seguro de que nos permitirá seguir navegando por moitos anos cara a un luminoso horizonte.

### Dirección:

Paula Raña Míguez

### Comunicación coa SGAPEIO:

[www.sgapeio.es](http://www.sgapeio.es)  
[secretaria@sgapeio.es](mailto:secretaria@sgapeio.es)  
[boletin@sgapeio.es](mailto:boletin@sgapeio.es)  
Twitter @sgapeio  
LinkedIn Sgapeio

### Depósito Legal:

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

# #SGAPEIO

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

Facultade de Matemáticas

Rúa Lope Gómez de Marzoa s/n Campus Vida-15782 Santiago de Compostela (A Coruña)

<http://www.sgapeio.es/>



# Algoritmos no combate de incendios forestais

por Ana Belén Buide Carballosa, Manuel Antonio Novo Pérez, María José Ginzo Villamayor

**A**na e Manuel son graduados en Matemáticas pola USC. Actualmente, son investigadores contratados no Instituto Tecnolóxico de Matemática Industrial (ITMATI).

María é licenciada en Matemáticas coa especialidade en Estatística e Investigación Operativa tamén pola USC. É investigadora contratada na USC e colaboradora de ITMATI.

Os tres son egresados do Máster interuniversitario en Técnicas Estatísticas.

## O proxecto ENJAMBRE

España é un dos países europeos con maior superficie forestal, pois conta cun total de 27,7 millóns de hectáreas de superficie forestal, o que equivale a máis da metade do seu territorio. Á súa vez, tamén é un dos países máis afectados por incendios forestais, aparecendo nos primeiros postos en número de incendios e superficie afectada. En concreto, o noroeste de España, e especialmente Galicia, é unha das zonas máis afectadas nos últimos anos.

Debido a esta problemática nacional e global xorden varios proxectos de investigación e de transferencia. A nivel de transferencia, a través do Centro para o Desenvolvemento Tecnolóxico Industrial (CDTI) téñense subvencionado diferentes proxectos, moitos deles a modo de consorcio formado por empresas, organismos públicos de investigación (OPIs) e outras institucións, ao abeiro de diversos programas. Por exemplo, a nivel internacional, o proxecto "Sistema intelixente para detectar incendios forestais e dirixir a extinción de incendios baseado en visión artificial, mapas 3D e tecnoloxía GIS" (EFIRE) do programa Eurostars (período de execución 2011 a 2014). A nivel nacional, o proxecto "Tecnoloxías para o combate integral contra incendios forestais e para a conservación dos nosos bosques" (PROMETEO), que contou co maior financiamento concedido ata a data, do Programa CENIT (período de execución 2010 a 2014). No caso de Galicia o proxecto "Tecnoloxías avanzadas para a extinción de grandes incendios forestais" (LUMES), do programa INNTERCONECTA (período de execución 2013 a 2014), e o proxecto "Misións críticas de emerxencias con medios aéreos tripulados e non tripulados en voo cooperativo" (ENJAMBRE) do programa CIEN (período de execución 2014 a 2018). O equipo de investigación do que forman parte os autores deste traballo participou no consorcio dos dous últimos a través do Instituto Tecnolóxico de

Matemática Industrial (ITMATI). A continuación preséntase con máis detalle a participación en ENJAMBRE e a concreción en dous dos algoritmos o "Algoritmo de Rutas de Escape das Brigadas" e o "Algoritmo de Eficiencia nas Descargas de Auga".

O proxecto ENJAMBRE céntrase en tres prioridades da Estratexia de Seguridade Nacional: a ordenación de fluxos migratorios, a protección ante emerxencias e catástrofes e a seguridade marítima, fronte as cales se realizan operacións de observación e de intervención para manter a seguridade. Os incendios forestais englobáanse na prioridade de protección ante emerxencias e catástrofes, e o principal obxectivo neste contexto é o desenvolvemento de tecnoloxías, entre as que se atopa un servizo de misións críticas de emerxencias que permita a actuación, de forma cooperativa e segura, de aeronaves tripuladas e non tripuladas nun mesmo espazo aéreo, desempeñando estas últimas tarefas de observación ao servizo de aeronaves tripuladas de intervención, facilitando a toma de decisións durante as operacións e reducindo os custos das mesmas.

Como xa se comentou, o proxecto está organizado en forma de consorcio, liderado pola multinacional Babcock, e conta coa participación de empresas como COREMAIN, PILDOLABS, ALAVA e TEMAI e as OPIs Centro Avanzado de Tecnoloxías Aeroespaciales (FADA-CATEC), ITMATI, Universidad de Alicante (UA), Universidad de Córdoba (UCO), Universidad Politécnica de Cataluña (UPC) e Universidad de León (ULE).

A participación de ITMATI céntrase no desenvolvemento de algoritmos matemáticos que axuden na toma de decisións na extinción de incendios forestais. Os principais son os seguintes: **estimación do perímetro dun incendio** a través de imaxes termográficas empregando, entre outras, técnicas de estimación de conxuntos; **algoritmo de anticolisión** entre as aeronaves que operan nun incendio a partir dos datos en tempo real das súas monitorizacións, que nos informan da velocidade e da traxectoria que seguen; **algoritmo de rutas de escape para as brigadas** que operan na extinción tendo en conta a orografía do terreo, a vexetación e vías existentes; **algoritmo de eficiencia nas descargas de auga** realizadas polos medios aéreos que traballan no proceso de extinción e **algoritmo de xestión de recursos aéreos** tendo en conta a normativa vixente para pilotos e aeronaves.

## Algoritmo de Rutas de Escape das Brigadas

Na extinción dun incendio forestal o uso de brigadas é un elemento esencial para o control do incendio desde a terra. Por isto, é necesario manter unha boa comunicación e organización das mesmas para que se enfronten con seguridade a este, evitando que o incendio, na súa evolución, chegue a cercalas. Esta situación pode estar motivada porque o terreo situado por detrás das brigadas se vai secando polo efecto da enerxía que desprende o incendio, de maneira que aumenta a probabilidade de que se incendie e, como consecuencia, a brigada quede atrapada, privándoa dunha ruta de escape.

O obxectivo deste algoritmo é calcular unha ruta de escape para a evacuación dos medios de extinción terrestres, proporcionándolle ao director de extinción unha ruta segura para retirarse, desde o incendio, ata o destino seleccionado. Para isto empréganse funcións de tempo mínimo en base ás distintas velocidades de desprazamento dunha persoa en función da pendente do terreo, do tipo de vexetación e do tipo de vía. As rutas obtidas conectan a posición das brigadas de extinción coa zona de seguridade máis próxima, evitando os posibles obstáculos que van encontrar no camiño e non poden sortear.

Para a obtención da **ruta de escape** para a brigada, dende a posición que se atopa ata unha zona de evacuación segura dada de antemán, é preciso calcular o **custo de desprazamento** da brigada polo terreo, isto é, a velocidade máxima á que se pode desprazar en función da pendente do terreo, da vexetación existente, das vías dispoñibles e dos obstáculos presentes. A metodoloxía empregada para realizar este cálculo foi a seguinte:

1. A partir do modelo dixital de elevacións (MDE), do Instituto Geográfico Nacional (IGN), que representa as elevacións sobre o terreo, obtense a pendente do mesmo, segundo Horn (1981).
2. Unha vez calculada a pendente, obtense a velocidade máxima de desprazamento aplicando a fórmula proposta por Tobler (1993).
3. Reclasifícase a vexetación existente, asignando uns coeficientes que reducen a velocidade en función da estrutura da vexetación, baseados nos valores propostos por Wood and Schmidtlein (2011) para cada tipo de modelo de combustible, de acordo coa clasificación Prometheus (Arroyo et al., 2006).
4. Cálculanse os coeficientes que reducen a velocidade en función do tipo de vía baseados nos valores propostos por Wood and Schmidtlein (2011). Para isto, previamente, foi preciso obter unha capa vectorial cos distintos tipos de vías considerados (estrada convencional, camiño, pista, carreiro, cortalumes). Fíxose uso da base de datos xeográfica a escala 1:25000 dispoñible para todo o territorio nacional (BTN25) do IGN, que dispón dun amplo contido de fenómenos xeográficos en formato

SHAPEFILE. Dos datos correspondentes á zona do incendio extraéronse os elementos que conforman os distintos tipos de vías, facendo uso do código que presentan na súa nomenclatura, uníndoos todos nunha única capa vectorial cun novo campo co correspondente valor do coeficiente de redución da velocidade e que posteriormente foi preciso rasterizar.

5. Créase unha capa *ráster* onde cada píxel representa a velocidade máxima á cal se pode desprazar a brigada en función da pendente, da vexetación e do tipo de vía.
6. Extráese da base de datos BTN25 da zona do incendio os elementos considerados como obstáculos (ríos, encoros...) seguindo o procedemento indicado no punto anterior (5). Neste caso, dada a tipoloxía dos elementos considerados como obstáculos, obtéñense dúas capas vectoriais (unha, tipo liña e outra, polígono) que tamén se rasterizan.
7. Intégranse os obstáculos e o incendio na capa *ráster* de velocidade, asignándolle valores nulos aos píxeles correspondentes. O incendio vén dado por unha capa vectorial que define o perímetro do mesmo.

Calculada a capa *ráster* dos custos de desprazamento coa metodoloxía anterior, obtense a ruta de escape da brigada desde a posición na que se atopa ata unha zona segura seleccionada previamente. Para iso, transfórmase a capa *ráster* nunha matriz de transición establecendo o valor de transición entre cada par de píxeles á media dos dous valores que está conectando, e permitindo movementos diagonais entre os oito píxeles adxacentes. Tras isto, faise unha corrección consistente en dividir os valores da matriz entre a distancia dos centros dos píxeles, debido a que moverse en diagonal supón percorrer maior distancia que moverse en liña recta. Finalmente, calcúlase o camiño máis curto entre o punto de inicio e o punto de destino (van Etten, 2013).

A ruta obtida proporcióname como unha capa vectorial nos formatos SHAPEFILE e KML que se poden visualizar mediante ferramentas que empregan sistemas de información xeográfica (SIX) e os atributos dan conta da velocidade media da ruta, do tempo que levaría percorrela e da distancia da mesma. Tamén se proporcionan as coordenadas dos puntos característicos da ruta obtida, entendendo como característicos aqueles onde a traxectoria cambia, xunto co rumbo que se debe seguir para chegar ao seguinte punto e a distancia (en liña recta), seguindo ese rumbo, que hai que percorrer para alcanzar o seguinte punto. Isto resulta de utilidade, debido a que nunha situación extrema ou en determinadas localizacións o GPS pode perder o sinal, polo que os brigadistas non poderían guiarse pola ruta proporcionada, pero, ao dispoñer destes puntos e da información proporcionada,

coa utilización dun compás poderían chegar á zona segura. Na Figura 1 amósase o resultado dunha execución do algoritmo. En cor laranxa represéntase o incendio, o punto azul fai referencia á posición inicial da brigada e o verde, á zona segura cara a onde se quere desprazar a brigada.

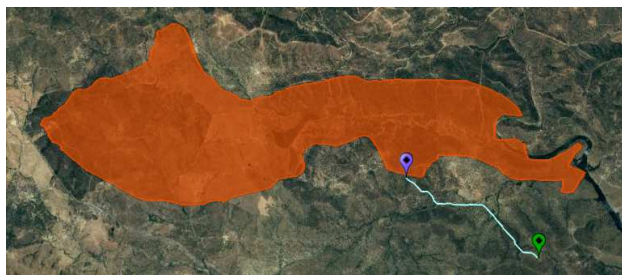


FIGURA 1: Resultado da execución do Algoritmo Rutas de Escape. En cor laranxa represéntase o incendio, o punto azul fai referencia á posición inicial da brigada e o verde, á zona segura á que se quere desprazar a brigada. En cor azul claro amósase a ruta de escape.

## Algoritmo de Eficiencia nas Descargas de Auga

Os medios aéreos xogan un papel moi importante na extinción dos incendios forestais. Por iso, considérase importante coñecer se as descargas de auga que estes realizan están sendo efectivas para saber se se están utilizando unha cantidade adecuada de recursos. Medir o anterior é o obxectivo deste algoritmo.

O **Algoritmo de Eficiencia nas Descargas de Auga** emprega datos da monitorización das aeronaves que participan nun incendio, as cales proporcionan a posición da aeronave e outras características do voo, por exemplo, o momento no que se realizan certos eventos, como pode ser o instante no que realizan as descargas de auga. Para o cálculo da **eficiencia das descargas** son precisos outros datos, non só as monitorizacións, como son imaxes térmicas do incendio, unha delas, anterior ás descargas de auga e outra, posterior, necesarias para coñecer a diferenza de temperatura entre ambas as dúas. Coa información anterior pódese xerar unha imaxe tipo *ráster* na que, en cada píxel, se proporciona a eficiencia mediante un indicador que consiste no cociente da diferenza de temperatura entre as dúas imaxes pola cantidade de auga descargada.

A **estimación da marca de auga**, isto é, a área cuberta polas descargas de auga realizadas, é necesaria para o cálculo da eficiencia. Esta área supónse que ten forma elíptica e calcúlase en base a modelos estatísticos de regresión (Rodríguez y Silva e González-Cabán, 2010), que permiten calcular o longo e o largo da elipse a partir do volume da descarga de auga, a altura da mesma e a

velocidade da aeronave, orientando a elipse en función da dirección da aeronave no momento da descarga e supoñendo que a auga da descarga se distribúe de forma homoxénea. Unha vez obtidas as marcas de auga de cada unha das descargas realizadas, calcúlase unha capa *ráster* que representa a marca de auga total e onde cada píxel representa o total de litros descargados polas aeronaves que actúan no incendio. Finalmente, calcúlase a capa de eficiencia das descargas a partir das elipses estimadas e das dúas imaxes térmicas, proporcionando para cada píxel a diminución de temperatura por litro de auga descargado polas aeronaves.

Cando se utiliza o algoritmo en tempo real, é común non contar con imaxes térmicas do incendio, xa que, a toma de imaxes actualmente é unha das misións da aeronave de coordinación que non ten por que estar presente durante toda a duración do incendio. A pesar diso, segue sendo de interese coñecer a marca de auga das descargas que realizan as aeronaves para determinar se se están a realizar nos lugares axeitados. Por iso, considerouse esta opción no algoritmo, e, no caso de non dispoñer de imaxes térmicas do incendio entre as que se están a realizar as descargas, proporciónase unha imaxe *ráster* xeoreferenciada da última descarga realizada por cada aeronave, o que permite coñecer onde se están realizando as descargas.

Na Figura 2 represéntanse as marcas de auga das descargas realizadas por varias aeronaves que participaron nun incendio en Galicia no verán do 2017.

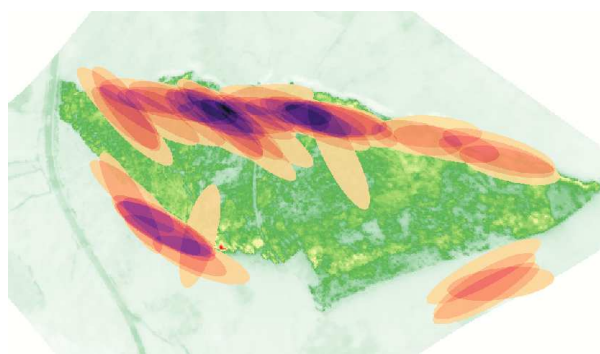


FIGURA 2: Marca de auga das descargas realizadas por varias aeronaves que participaron na extinción dun incendio ocorrido en Galicia no verán do 2017.

Para o desenvolvemento destes dous algoritmos empregouse o software estatístico R, <https://www.r-project.org/> e QGIS <https://qgis.org/es/site/>.

## Agradecementos

Os autores desde traballo agradecen o apoio do proxecto ENJAMBRE do CDTI con expediente IDI-2014 1237 e aos fondos FEDER a través do Programa Operativo de Crecimiento Inteligente. Igualmente, a todo o persoal humano do mesmo e, de maneira especial, aos investigadores de ITMATI que colaboraron no mesmo.

# O paquete `fda.usc`

Manuel Oviedo de la Fuente

A análise dos datos funcionais (FDA no seu acrónimo en inglés) é unha área de investigación moi activa nos últimos anos e aparece naturalmente na maioría das áreas científicas: econometría (curvas diarias de demanda ou prezo da electricidade), medio ambiente (curvas de niveis de contaminantes), medicina (curvas de crecemento), quimiometría (datos de espectrometría), etc.

O paquete `fda.usc` (dispoñible nesta [ligazón](#)) presenta un marco común para o software R que aglutina de maneira homoxénea e estruturada as propostas das dúas tendencias principais na análise de datos funcionais; o libro de Ramsay e Silverman (2005), a partir dunha perspectiva paramétrica (limitada a espazos de Hilbert), e o libro de Ferraty e Vieu (2006) dende o punto de vista non paramétrico (espazos funcionais métricos ou semi-métricos).

Febrero-Bande e Oviedo de la Fuente (2012), presentaron unha revisión detallada do estado da arte relacionado co FDA e os métodos estatísticos que poden aplicarse a un conxunto de datos funcionais. Dende aquela, fixéronse 21 versións do paquete co obxectivo de manter, incorporar e propoñer os procedementos máis relevantes publicados na literatura do FDA, algúns deles recollidos en Oviedo de la Fuente (2019). A Figura 1 mostra a evolución das descargas do paquete.

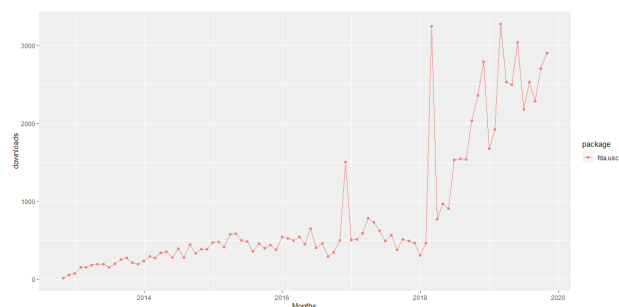


FIGURA 1: Evolución das descargas do paquete `fda.usc`.

O `fda.usc` ten incorporados os seguintes procedementos que seguen os estándares de programación de R:

- A representación do dato funcional empregando a clase “`fdata`”: As funcións incluídas nesta sección permiten definir, operar, transformar, manipular e representar un conxunto de datos funcional de moitas formas, incluíndo derivadas, representación en bases e métodos do núcleo. Para o conxunto

de datos da AEMET, o panel superior esquerdo da Figura 2 mostra a curva de temperatura rexistrada (en negro), a suavizada empregando B-splines (en vermello) e polo núcleo gaussiano (en verde).

- O computo de medidas de centralidade e periferia empregando a profundidade funcional (vexa axuda co comando `help(Depth)` en R). O panel superior dereito da Figura 2 mostra as medidas (localización media) e dispersión (banda de confianza *bootstrap* do 95%), rexistradas para os niveis de  $NO_x$  nos días laborables ou non laborables.
- Modelos de regresión funcional (funcións co prefixo “`fregre.`”): lineal, parcialmente-lineal, xeneralizado, aditivo, resposta funcional e erros dependentes (Oviedo de la Fuente et al. 2018).
- Clasificación supervisada (funcións co prefixo “`classif.`”). Os paneis na parte inferior da Figura 2 mostra: na esquerda, a segunda derivada dunas curvas espectrométricas (coloreadas polo contido de graxa), e á dereita, o resultado do clasificador  $DD^G$  empregando a profundidade modal (Cuesta et al., 2017).
- Proba de bondade do axuste do modelo lineal funcional con resposta escalar (funcións co prefixo “`f1m.`”), Garcia-Portugues et al., (2014).
- Utilidades para a detección atípica (funcións co prefixo “`outliers.`”), ANOVA funcional (funcións co prefixo “`anova.`”), clustering funcional (función `kmeans.fd`) e selección de variables na regresión (función `fregre.gsam.vs`), por citar as máis destacadas.

O paquete foi incluído no conxunto dos 7 paquetes recomendados no *Task View* no campo do FDA <https://cran.r-project.org/web/views/FunctionalData.html>.

Para máis documentación sobre o uso do paquete consulte a súa axuda na seguinte [ligazón](#) e a tarxeta de referencia:

<http://doi.org/10.5281/zenodo.3386752>.

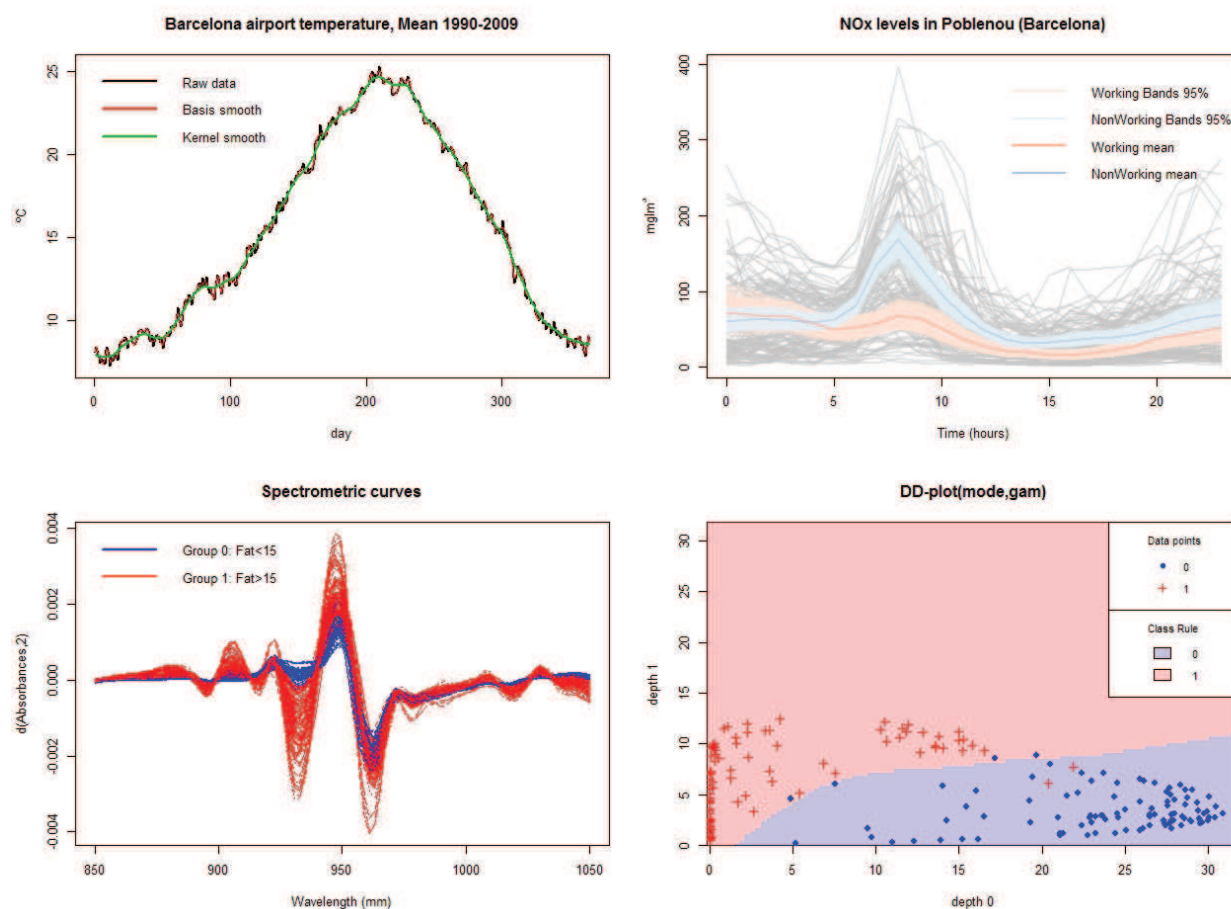


FIGURA 2: Exemplos gráficos empregando ferramentas do `fda.usc`.

### Agradecimentos:

Grazas ao co-autor do pacote Manuel Febrero Bande e aos contribuintes: Pedro Galeano, Alicia Nieto e Eduardo García-Portugués.

### Financiación:

Proxecto MTM2016-76969-P do Ministerio de Economía e Competitividad da Axencia Estatal de Investigación.

Cuesta-Albertos, J. A., Febrero-Bande, M., and Oviedo de la Fuente, M. (2017). The  $DD^G$ -classifier in the functional setting. *Test*, 26(1), 119-142.

Febrero-Bande, M., González-Manteiga, W., and Oviedo de la Fuente, M. (2019). Variable selection in functional additive regression models. *Computational Statistics*, 34(2), 469-487.

Febrero-Bande, M. and Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: the R package `fda.usc`. *Journal of Statistical Software*, 51(4):1-28.

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer.

García-Portugués, E., González-Manteiga, W. and Febrero-Bande, M. (2014). A goodness-of-fit test for the functional linear model with scalar response. *Journal of Computational and Graphical Statistics*, 23(3), 761-778.

Oviedo de la Fuente, M., Febrero-Bande, M., Muñoz, M. P., and Domínguez, A. (2018). Predicting seasonal influenza transmission using functional regression models with temporal dependence. *PLoS one*, 13(4), e0194250.

Oviedo de la Fuente, M. (2019). Advances in functional regression and classification models. Doctoral dissertation, <http://hdl.handle.net/10347/18236>.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer.



# Actividades da SGAPEIO



## XIV Congreso Galego de Estatística e Investigación de Operacións

por Javier Roca Pardiñas e Alejandro Saavedra Nieves

Como resposta á petición da Sociedade Galega para a Promoción da Estatística e da Investigación de Operacións (SGAPEIO), o Departamento de Estatística e Investigación Operativa da Universidade de Vigo (UVigo) organizou o XIV Congreso Galego de Estatística e Investigación de Operacións (SGAPEIO2019), celebrado na Escola de Enxeñaría Industrial (sede cidade) da UVigo do 24 ao 26 de outubro do 2019.

Este encontro forma parte dos congresos bianuais promovidos pola SGAPEIO e, coma en edicións anteriores, a súa finalidade foi a divulgación na investigación e no ensino das innovacións na Estatística e na Investigación de Operacións. Nesta edición, a programación oficial contou cunha sesión especial de Biometría, que foi organizada conxuntamente pola propia SGAPEIO e a Sociedade Portuguesa de Estatística (SPE), e que dá continuidade aos Encontros Galaico-Portugueses organizados polas dúas sociedades dende o ano 2013. Esta colaboración permite a promoción e divulgación da Estatística e da Investigación de Operacións en distintas áreas de aplicación, fomentando novas ideas e colaboracións de interese para ambas as dúas comunidades.

O congreso tivo un amplo programa científico composto por catro conferencias plenarias, dúas mesas redondas, sete sesións de comunicacións orais e dúas sesións de pósters. Como xa se mencionou, tivo lugar unha sesión especial de Biometría, organizada pola SGAPEIO e a SPE. Contouse cunha exposición do Museu de Matemàtiques de Catalunya e impartíronse dous obradoiros. Ademais, houbo un amplo número de actividades sociais enmarcadas na cidade de Vigo e na súa contorna.

O congreso contou cun total de 115 persoas inscritas, coa participación de representantes das tres universidades galegas e tamén de fóra de Galicia, desde a Universidad Complutense de Madrid ata a Universidad de la Laguna, a UNED, a Universidad Miguel Hernández de

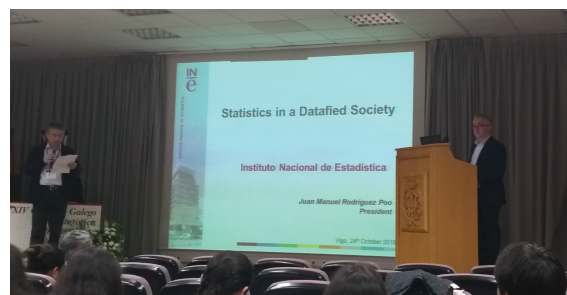
Elche, pasando tamén polas de Salamanca e Valencia. Do ámbito internacional, son salientables a Universidad Nacional de San Agustín de Perú, así como varias universidades portuguesas, de Bélxica, de Suíza e dos Países Baixos. Ademais de representantes do ámbito universitario, contouse tamén coa participación de numerosos profesionais do ensino medio, da propia Xunta de Galicia, dos Institutos, galego e nacional, de Estatística; do Centro Tecnolóxico do Mar, así como de empresas e organismos de carácter privado como Abanca ou Hijos de Rivera.

Presentáronse un total de 51 traballos, dos cales 33 foron presentacións orais e 18, comunicacións en formato póster.

A continuación, detállase de maneira resumida o cronograma das actividades científicas e sociais acontecidas durante os tres días do congreso.

### Xoves, 24 de outubro

O encontro iniciouse co acto oficial de apertura do Congreso. Contou coa presenza da Tenente de Alcalde e Concelleira delegada da área de Urbanismo do Concello de Vigo, María Xosé Caride, en representación do alcalde da cidade; do Deputado Provincial Carlos López Font, en representación da presidenta da Deputación de Pontevedra; do director do Instituto Galego de Estatística, José Antonio Campo Andión e do presidente do comité organizador do SGAPEIO2019, Javier Roca Pardiñas. Unha vez rematada a apertura oficial do congreso, tivo lugar a primeira sesión plenaria a cargo do presidente do Instituto Nacional de Estadística, Juan Manuel Rodríguez Poo.



Tras un breve receso, celebrouse a segunda das sesións plenarias programadas, impartida por Juan José Salazar González, profesor da Universidade de La Laguna.



Ao remate discutiuse sobre a “Investigación Operativa e Empresa” durante a primeira mesa redonda. Esta foi moderada por Salvador Naya Fernández (Universidade da Coruña, UDC), e contou coa presenza de Silvia Lorenzo Freire (UDC), Julio González Díaz (Universidade de Santiago de Compostela, USC) e Jorge López Muñiz (Grupo Hijos de Rivera - Estrella Galicia) como poñentes.

Como feito relevante, cómpre mencionar a Sesión especial de Biometría que tivo lugar á primeira hora da tarde do xoves. Foi moderada por María José Ginzo Villamayor (USC) e nela participaron Rosa Crujeiras (USC), Magda Monteiro (Universidade de Aveiro) e Soraia Pereira (Universidade de Lisboa). En paralelo coa anterior, presentáronse 11 comunicacións orais noutras dúas sesións: “Investigación operativa e teoría de xogos”, e “Estatística pública”.



A continuación expuxéronse os pósters correspondentes da primeira sesión e rematouse coa presentación de 12 comunicacións orais en tres sesións paralelas: “Aplicacións da estatística”, “Contribucións en bioestatística” e “Premios da SGAPEIO (I)”. A última das mencionadas correspóndese coa primeira sesión adicada aos premios ás mellores comunicacións presentadas no congreso SGAPEIO2019 por investigadores/as novos/as. Como

en convocatorias anteriores, ditos premios foron convocados en dúas categorías: á mellor comunicación teórica e á mellor comunicación aplicada.



Rematouse o día coa recepción oficial do congreso, ofrecida polo Concello de Vigo e realizada no Pazo de Quiñones de León, dentro da contorna do Parque de Castrelos (Vigo). Antes do discurso das autoridades presentes, contouse coa interpretación das Sete Cantigas de Amigo de Martín Códax (Brais González) por parte do Coro de Cámara Rías Baixas.

## Venres, 25 de outubro

O segunda día empezou coa celebración de dúas sesións de comunicación orais en paralelo, “Regresión” e “Premios da SGAPEIO (II)”, nas que se presentaron un total de 9 traballos.



A continuación celebráronse dúas sesións plenarias impartidas respectivamente por Carmen Armero, da Universitat de València, e por Ingrid Van Keilegom, da KU Leuven (Bélxica).



Entre ambas as dúas plenarios tivo lugar a segunda das sesións de pósters programadas no congreso. A actividade da mañá do venres concluíu coa celebración da Asemblea Xeral da SGAPEIO.

Pola tarde, e ao remate das actividades científicas programadas para ese día, os participantes desfrutaron dunha ruta en barco pola Ría de Vigo. Despois dunha breve travesía, puideron visitar as illas de San Simón e San Antón (Redondela).



A xornada concluíu co concerto do pianista vigués Brais González no Vitruvia Café.

### Sábado, 26 de outubro

Durante a mañá do último día impartíronse os dous obradoiros incluídos na programación do congreso:

- Minería de datos con R. Este taller, impartido por M. Carmen Iglesias Pérez, Ricardo Luaces Pazos e Antonio Vaamonde Liste da UVigo, concluíu cunha última sesión presencial tras máis de 30 horas impartidas en liña.
- Elaboración de gráficos con R. Este taller foi impartido por Javier Tarrío Saavedra, da UDC.

Unha vez finalizados, houbo un taller da exposición do Museu de Matemàtiques de Catalunya. Foi impartido polos membros do museo que nos acompañaron durante os tres días de congreso, Enric Brasó e Aleix Mestre.



Tras este pequeno taller, os asistentes puideron asistir á mesa redonda "A importancia da Estatística na vida diaria" e que foi moderada por Covadonga Rodríguez-Moldes Rey (IES de Mugardos). Nela interviñeron Jesús Díaz López (IES Universidade Laboral de Culleredo), José Roberto García Olano (Instituto Galego de Estatística) e Xaime Suárez Blanco.

Antes do peche do congreso, tivo lugar o acto de entrega de premios aos investigadores novos. Foron os traballos de Marta Cousido Rocha e Alejandro Saavedra Nieves os que resultaron premiados nas categorías á mellor comunicación teórica e á mellor comunicación aplicada, respectivamente. Despois deste momento, o presidente da SGAPEIO, César Sánchez Sellero, xunto con Jacobo de Uña Álvarez e Javier Roca Pardiñas, presidentes do comité científico e do organizador do congreso respectivamente, pecharon o congreso.



Finalmente, celebrouse o xantar de confraternidade do congreso no Restaurante Albatros, ubicado na Estación Marítima de Vigo.

Gustaríanos agradecer á Universidade de Vigo, ao Concello de Vigo, á Deputación de Pontevedra, á Xunta de Galicia, aos principais grupos de investigación de estatística de Galicia (SiDOR, ECOSOT, MODESTYA, MODES), ECOBAS, rede Biostatnet, e a outras entidades, tanto públicas como privadas, que de maneira desinteresada colaboraron na realización deste evento. Este agradecemento é extensible tamén a todos aqueles asistentes ao congreso que, coa súa presenza, o fixeron posible.

Para rematar, non podemos esquecernos dos nosos compañeiros e das nosas compañeiras do comité organizador: Javier Roca Pardiñas - Presidente (UVigo), Ana María Domínguez Punte (IES Castelao - Vigo), María José Ginzo Villamayor (USC), María Gómez Rúa (UVigo), María del Carmen Iglesias Pérez (UVigo), María Jesús Longa Sayáns (UVigo), Leticia Lorenzo Picado, Juan Carlos Pardo Fernández (UVigo), Alejandro Saavedra Nieves (UVigo) e Antonio Vaamonde Liste (UVigo). O gran traballo realizado por este equipo fixo posible que o congreso resultase todo un éxito.



### Sesión de biometría no XIV Congreso Galego de Estatística e Investigación de Operacións

por Rosa Crujeiras Casais

A Sociedade Portuguesa de Estatística (SPE) e a SGAPEIO organizaron conxuntamente unha sesión de biometría o venres 25 de outubro, no marco do XIV Congreso Galego de Estatística e Investigación de Operacións. Magda Monteiro, da Universidade de Aveiro, e Soraia Pereira, da Universidade de Lisboa, foron as representantes da SPE, participando Rosa M. Crujeiras (da Universidade de Santiago de Compostela) por parte da SGAPEIO.

Na sesión presentáronse problemas de aplicación práctica no ámbito da bioloxía e do medio ambiente, abordados con diferentes metodoloxías. Magda Monteiro centrou a súa presentación na detección de puntos de cambios nas taxas de crecemento de series de temperaturas en cidades europeas, como un indicador do cambio climático. Pola súa banda, Soraia Pereira analizou, mediante ferramentas da estatística Bayesiana (concretamente, utilizando INLA), o patrón xeostatístico dos ovos de sardiña na costa de Portugal, un estudo cuxos resultados poden contribuír a unha mellora da xestión dos bancos de peixes con obxectivos de sustentabilidade económica e ambien-

tal. Finalmente, Rosa M. Crujeiras presentou modelos non paramétricos de regresión con covariables categóricas na presenza de observacións de natureza circular, unha metodoloxía con aplicación no estudo de orientacións de animais. O exemplo concreto que se analizou trataba de modelar a orientación das pulgas de praia en distintos escenarios (que trataban de recoller a intervención na paisaxe), tendo en conta a influencia da temperatura ou do acimut do sol.



A novidade e aplicabilidade das temáticas tratadas, así como a discusión sobre as presentacións, puxeron de manifesto o importante pulo que se lle ten dado á estatística aplicada ao ámbito biométrico en Galicia e Portugal: un campo onde as dúas comunidades científicas teñen moito que achegar, aprender e compartir facendo posible a transferencia de coñecemento á sociedade en ámbitos tan relevantes como o control medioambiental ou a conservación das especies.

A sesión foi organizada por María José Ginzo, secretaria da SGAPEIO, e Giovani Silva.

## Asemblea Xeral da SGAPEIO o día 25 de outubro de 2019

por M<sup>a</sup> José Ginzo e Paula Raña

O día 25 de outubro celebrouse no Salón de Actos da Escola de Enxeñería Industrial (sede cidade) da Universidade de Vigo unha Asemblea Xeral de SGAPEIO.



En primeiro lugar aprobouse a Acta da Asemblea Xeral do 13/04/2018. A continuación informouse do estado do número de socios, que actualmente é de 175 socios individuais e de 5 institucionais. Animouse aos socios a fomentar a incorporación de novos socios. En canto ao *Informest*, coméntase que se segue a traballar cos 3 números habituais por ano. Anímase a todos os socios que queiran participar a que se poñan en contacto coa sociedade.

A continuación, fíxose un repaso polas actividades previstas en 2019 e 2020. Infórmase de que o día 4 de outubro se celebrou na Facultade de Matemáticas o *PhDay* ([web](#)) entre os alumnos de doutoramento das tres universidades galegas. A SGAPEIO outorgou o premio á mellor comunicación oral e agasallou a todos os asistentes cun libro de Estatística en Caricaturas.

O día 24 de outubro tivo lugar a sesión de biometría coorganizada coa Sociedade Portuguesa de Estatística (SPE). Os convidados foron: Rosa Crujeiras Casais (USC, convidada da SGAPEIO), Magda Monteiro (Universidade de Aveiro, convidada da SPE) e Soraia Pereira (Universidade de Lisboa, convidada da SPE).

Infórmase que dende a SGAPEIO se organiza a Xornada de Formación para Profesores de Ensino Secundario en colaboración con AGAPEMA. Participarán os profesores dos traballos gañadores na IX edición do Concurso de “*Incubadora de Sondaxes e Experimentos*”, M<sup>a</sup> José Ginzo Villamayor e María Martín Vila impartirán un obradoiro e Paula Raña Míguez impartirá a conferencia divulgativa ([web](#)).

Así mesmo infórmase de que o día 8 de novembro terá lugar a sesión de biometría coorganizada coa SPE e a Sociedade Española de Biometría (SEB) no Congreso Nacional da SPE en Amarante. Os convidados serán: Paula Saavedra Nieves (USC, convidada da SGAPEIO), Andrea Meilán Vila (UDC, convidada da SGAPEIO), Nuno Sepúlveda (London School of Hygiene and Tropical Medicine convidado da SPE) e Juan Ramón González (Universidad Autónoma de Barcelona, convidado da SEB).

En canto ao concurso “*Incubadora de Sondaxes e Experimentos*”, explícanse os pasos que se levan dado na organización e posta en marcha da súa décima edición. Infórmase de que na fase nacional modificaron un pouco as bases e que a SGAPEIO, na medida das súas posibilidades, tratará de adaptarse ás mesmas.

Sobre o XIV Congreso Galego de Estatística e Investigación de Operacións, Javier Roca Pardiñas comenta que está contento do transcurso do que vai de Congreso e quere agradecer o traballo realizado tanto polo comité científico como polo comité organizador.

César Sánchez comenta que asistiu á xornada de clausura dos *Estalmat*. Entregouse a cada un dos egresados un exemplar do libro Estatística en Caricaturas como agasallo.

En canto ao vindeiro congreso (XV Congreso Galego de Estatística e Investigación Operativa), infórmase de que, de momento, non se pode anunciar onde será, pero que por quenda, en ausencia doutras propostas, correspondería en Santiago de Compostela. O Comité Executivo está aberto a suxestións e nos vindeiros días reunirse e decidirá quen será o próximo organizador e cal será a cidade onde se celebrará.

Infórmase de que se asinou recentemente un convenio de reciprocidade coa Sociedade Portuguesa de Estatística (SPE).

A profesora Rosa M. Crujeiras Casais, como organizadora do Congreso *Advances in Directional Statistics (ADISTA20)*, quere solicitar colaboración da SGAPEIO para financiar o premio “Kanti Mardia” ao mellor póster.

Para rematar, M<sup>a</sup> José Ginzo Villamayor explica o balance para o ano 2018 e o orzamento previsto para o ano 2019. Apróbanse ambos por asentimento.



## Novidades na información estatística publicada

### Escenario macroeconómico

O [Escenario macroeconómico](#) refírese á evolución prevista do Produto Interior Bruto (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.

O método utilizado é un modelo de estimación de agregados macroeconómicos a curto prazo, nun marco coherente cos escenarios propostos para as economías española e internacional. Un resumo do mesmo pode verse na seguinte [ligazón](#).

Segundo o Escenario macroeconómico, a economía de Galicia prevese que medre un 2 % no ano 2019 e un 1,9 % no ano 2020. Por outra parte, a taxa de paro sitúase no 11,4 % no 2019 e no 10,6 % no 2020.

### Carga de poboación estacional dos concellos de Galicia

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.

A carga de poboación representa a media trimestral de persoas equivalentes a tempo completo que soporta un concello. Obtense a partir da poboación residente, sumándolle a poboación de entrada (non residentes presentes) e restándolle a poboación de saída (residentes ausentes). O saldo poboacional entre os non residentes que entran e os residentes que saen mide a diferenza entre a poboación presente e a poboación residente.

No primeiro, segundo e cuarto trimestre de 2017, os concellos con maiores saldos poboacionais son as sete grandes cidades, xunto cos concellos de Sanxenxo, O Porriño, O Grove (no segundo trimestre) e San Cibrao das Viñas (no primeiro e cuarto trimestre). Agora ben, no terceiro

trimestre de 2017, os concellos de Ferrol, O Porriño e San Cibrao das Viñas xa non se atopan entre os dez que teñen un maior saldo poboacional, sendo substituídos polos concellos costeiros do Grove, Nigrán e Baiona.

### Enquisa estrutural a fogares. Coñecemento e uso do galego

A Enquisa estrutural a fogares (EEF) é unha actividade estatística anual dirixida aos fogares galegos co obxecto de obter información sobre 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 2018 un dos módulos específicos está dedicado ao coñecemento e uso da lingua galega. Neste módulo analízanse, entre outras cousas, o grao de coñecemento do idioma galego nas súas distintas formas (oral e escrito) e o seu uso en distintos ámbitos sociais.

No ano 2018 o 57,59 % dos galegos e das galegas de 5 ou máis anos manifesta que sabe falar o galego moito; o 30,46 %, bastante e o 11,95 %, pouco ou nada. Entre a poboación menor de 15 anos, a porcentaxe de persoas que saben falar pouco ou nada o galego sitúase no 23,90 %, mentres que entre a poboación de 65 ou máis anos sitúase no 8,84 %.

Tamén en 2018, o 30,57 % da poboación de Galicia fala sempre en galego. Esta porcentaxe é inferior á rexistrada hai 5 anos en 0,63 puntos.

### Outras novidades

#### Pleno do Consello Galego de Estatística 21/06/2019

O 21/6/2019 celebrouse o Pleno do Consello Galego de Estatística no que se aprobou a acta da sesión anterior e se presentou o informe de seguimento do Programa Estatístico Anual 2018 (PEA 2018). O Director do IGE resumiu os resultados do PEA 2018, formado por 144 estatísticas propias e 33 actuacións complementarias de difusión e análise. O grao de execución global foi do 98 % e os indicadores de calendario amosan un 77 % de puntualidade e desfases entre datas de referencia e difusión de 13 meses, nas estatísticas estruturais, e de 3, nas conxunturais. Ademais, tamén se someteu á consideración do Consello unha addenda do proxecto técnico da operación Indicadores demográficos.

### Convenios

O 22 de outubro de 2019 firmouse o convenio de colaboración entre a Universidade da Coruña e o IGE para a realización de prácticas académicas externas de estudantes da Facultade de Socioloxía.

# Traballos de Estatística e Investigación Operativa no SUG

## Teses

*Contributions to distributional regression models. Applications in biomedicine*

**Jenifer Espasandín Domínguez.**

**Directores: Carmen Cadarso-Suárez, Francisco Gudel e Thomas Kneib.**

**Data da defensa: 19 de xullo de 2019. USC.**

O traballo de investigación realizado na presente tese parte da converxencia das liñas de investigación de dous grupos: un, de estatística (Group in Biostatistics and Biomedical Data Science, GI-2127) e outro, de medicina clínica (Research methods, C017), ambos pertencentes ao Instituto de Investigación Sanitaria de Santiago de Compostela e coordinados dende hai anos para levar a cabo investigación interdisciplinar no ámbito da Bioestatística.

Nesta tese, realízanse contribucións metodolóxicas no ámbito da Estatística, para a investigación de factores determinantes na glicación de proteínas. Especificamente, como resultado de varios proxectos previos, dispónse dunha ampla mostra da poboación xeral adulta que permite investigar diversos retos actuais no campo das enfermidades crónicas, como a diabetes. Ademais, contouse cos resultados da monitorización continua da glucosa intersticial durante unha semana. Dispónse, polo tanto, dos perfís de glucosa de cada individuo como dato funcional.

Especificamente, dende un punto de vista estatístico, nesta investigación revísanse e compáranse modelos de regresión distribucional frecuentistas e Baesianos para respostas univariantes e bivariantes, con gran flexibilidade tanto no modelado da variable resposta como nas posibles covariables predictoras, fronte aos modelos de regresión clásica.

Así mesmo, as ferramentas para a análise de datos funcionais son de grande utilidade para poder incorporar, nos modelos de regresión, toda a información dispoñible nos perfís de glucosa mencionados. Na presente investigación, adaptáronse as técnicas introducidas por Brockhaus et al. (2018) para permitir a incorporación de datos funcionais como covariables no ámbito da regresión distribucional univariante. No contexto multivariante, estendeuse a metodoloxía de McLean et al. (2014) para posibilitar a consideración de covariables funcionais nos modelos de regresión distribucional baseados en cópulas propostos por Marra and Radice (2017). Segundo o noso coñecemento, ata a data, non existe ningún outro modelo de regresión de cópula que permita modelar covariables funcionais cunha flexibilidade similar.

## Traballos Fin de Máster

Traballos do Máster Interuniversitario en Técnicas Estatísticas.

*Métodos estadísticos para a estimación da área de actividade dun animal: home range*

**Andrea Vilar Álvarez.**

**Directora: Beatriz Pateiro López.**

**USC, modalidade académica.**

O termo *home range* emprégase para designar a área de actividade dun animal, é dicir, a porción de terreo na que un animal realiza as actividades do día a día: alimentación, apareamento, coidado de crías. . .

Dende o punto de vista biolóxico, coñecer o *home range* dunha especie supón coñecerla máis profundamente,

o que facilita a xestión de recursos e, en xeral, a anticipación a situacións adversas, polo que é un concepto cun enorme interese biolóxico.

Así, neste traballo revisamos as técnicas existentes na literatura para a estimación do *home range* tanto con datos independentes como para datos con dependencia temporal, dende un punto de vista xeométrico que soamente nos dá unha estimación da área do *home range*, pero tamén dende un enfoque probabilístico, que ademais nos permite obter a frecuencia de uso. Logo, mediante un estudo de simulación avaliamos a calidade de cada estimador para detectar a xeometría subxacente do *home range* ou a distribución non uniforme das observacións. Finalmente, realízase a estimación do *home range* sobre un conxunto de datos reais.

O problema de empaquetado “The bin packing problema”

**Diego Frade Amil.**

**Directora: María Luísa Carpenle Rodríguez.**  
**USC, modalidade académica.**

O “bin packing problema” (BPP) ou problema de empaquetado en recipientes é un problema de optimización que consiste en asignar un conxunto de elementos cuns determinados tamaños ao mínimo número de recipientes respectando a capacidade destes. Este problema está presente en multitude de situacións, tanto da vida cotiá

como no ámbito industrial. Neste traballo realízase una revisión bibliográfica do mencionado problema, centrándose principalmente na versión unidimensional. Expóñense as súas principais características e algunhas variantes ou problemas relacionados, explícanse varios métodos de resolución e aplícanse a unha serie de conxuntos de datos e móstranse ferramentas ou recursos dispoñibles en internet para o tratamento deste problema (códigos de algoritmos, conxuntos de datos e aplicacións para visualizar as solucións). Ademais, realízase unha breve introdución ao BPP en máis dunha dimensión, incluíndo tamén algúns métodos para resolvelo e varios exemplos ilustrando cada caso.

## Traballos Fin de Grao

Traballos do Grao en Matemáticas presentados no curso 2018–2019.

Técnicas de clasificación en el contexto de “Big Data”

**Josefa Arán Paredes.**

**Titor: Wenceslao González Manteiga.**

O principal obxectivo deste traballo é introducir a análise discriminante e algunhas técnicas de clasificación, centrándose no seu papel no contexto de *Big Data*. No primeiro capítulo preséntanse os conceptos xerais da clasificación, as regras e a idea de “optimalidade”. Os principais métodos paramétricos de clasificación (regras lineais e cuadráticas) e algúns non paramétricos (regra dos  $k$  veciños máis cercanos e regras tipo kernel) son explicados no segundo capítulo. O terceiro trata sobre a avaliación das técnicas anteriormente definidas cando son aplicadas na práctica. Os dous últimos capítulos introducen os problemas que se poden achar nun contexto de *Big Data*, particularmente o caso en que a dimensión é maior que o número de observacións. Preséntanse dúas técnicas adaptadas a esta situación e finalmente ilústrase o seu funcionamento cun exercicio de simulación e cunha aplicación a unha base de datos real.

Problemas de rutas de vehículos e algoritmos de aforro: o algoritmo de Clarke and Wright

**Elena Blanco González.**

**Titora: Balbina V. Casas Méndez (cotitor: David Rodríguez Penas).**

Os Problemas de Rutas de Vehículos (VRP) son un exemplo de problemas para os que moitas veces non

é posible aplicar algoritmos que os resolvan en tempo polinomial, principalmente debido ao seu tamaño e á súa complexidade. Habitualmente para resolver exemplos deste tipo de problemas usáranse algoritmos heurísticos. Neste traballo presentamos o Problema do Viaxante de Comercio (TSP) para continuar cos Problemas de Rutas de Vehículos. A continuación enunciamos o algoritmo heurístico proposto por Clarke and Wright para o VRP con restricións de capacidade. Finalmente contamos cos datos dun problema de rutas da vida real que usamos para comparar as solucións obtidas usando unha implementación en R do algoritmo de Clarke and Wright coa solución obtida a dito problema mediante a súa modelación con linguaxe AMPL.

Asociación entre variables cualitativas: exemplo práctico de análise de correspondencias

**Jorge Castiñeiras Rella.**

**Titora: Beatriz Pateiro López.**

A análise de correspondencias é unha técnica estatística aplicada sobre táboas de continxencia que permite a visualización dunha nube de puntos multidimensional nun espazo de menor dimensión. Crea mapas da nube de puntos perdendo o mínimo de información posible para posteriormente realizar unha análise exploratoria dos datos rápida e eficaz. Busca asociación e relación entre variables en base a representacións que conserven ao máximo a varianza e posición orixinal dos datos. O desenvolvemento, explicación e implementación desta técnica van acompañados de casos e exemplos prácticos para a súa mellor comprensión. Finalmente desenvólvese un estudo médico completo buscando asociación entre diversas enfermidades, hábitos de vida e fortaleza de 820 pacientes.

## Problemas de regresión e clasificación usando SVM

**Ángela Dono Caramés.**  
**Titor: Manuel Febrero Bande.**

O obxectivo deste traballo é estudar un tipo de algoritmos de aprendizaxe coñecidos como *support vector machines*. Comezamos por introducir un caso sinxelo de clasificación binaria, o clasificador de marxe máxima, que empregamos para clasificar datos linealmente. Despois presentamos o concepto de función kernel, importante á hora de realizar clasificacións non lineais, e introducimos as variables de folgora para mellorar a capacidade de predición. A continuación, estudamos dous métodos que nos permiten traballar con problemas de clasificación multiclase. No segundo capítulo xeneralizamos as *support vector machines* aos problemas de regresión mediante o uso de funcións de perda  $\epsilon$ -insensitivas, que se basean en ignorar os erros, se estes son o suficientemente pequenos. Ao final do traballo, ilústrase a implementación en R destes algoritmos con dous exemplos reais.

## Aplicacións da teoría de xogos ao problema do *passepartout*

**Antón de la Fuente Suárez-Pumariega.**  
**Titora: Balbina V. Casas Méndez.**

Nun sistema de *passepartout*, un grupo de provedores de servizos deciden emitir un *passepartout* que permite aos seus clientes acceder aos seus servizos durante un período limitado de tempo e a un prezo fixo. O problema do *passepartout* consiste en como dividir os beneficios comúns obtidos pola venda dos *passepartout*. Neste traballo abordamos esta cuestión empregando diferentes ferramentas que nos proporciona a teoría de xogos cooperativos, centrándonos no caso no que o sistema de *passepartout* está formado exclusivamente por museos.

## Estabilización da varianza

**Alma García Pérez.**  
**Titor: César A. Sánchez Sello.**

Cando se constrúen intervalos de confianza para un parámetro, é frecuente enfrontarse ao problema de que o erro típico do estimador dependa do devandito parámetro. Ocorre, por exemplo, para unha proporción, para o coeficiente de correlación, ou para a media dunha distribución exponencial ou de Poisson. Para superar esta dificultade estúdanse dous métodos: o que denominamos *usual*, que se basea na estimación do erro típico, e o da *transformación estabilizadora da varianza*, que consiste en efectuar unha operación sobre o estimador que faga que o seu erro típico xa non dependa do parámetro descoñecido. Ademais, o estimador transformado segue disfrutando dunha aproximación normal, como consecuencia do método delta. Neste traballo fíxose un estudo de simulación en R, aplicado aos exemplos mencionados, para comparar os intervalos de confianza obtidos polo método usual e coa transformación estabilizadora da varianza. Obtivéronse mellores coberturas coa estabilización da varianza, en especial en distribucións continuas, xunto cunha mellor aproximación pola distribución normal.

## Modelos lineais de regresión cuantil

**Lucía Gil Rial.**  
**Titora: Mercedes Conde Amboage.**

Os modelos de regresión permiten describir o efecto de certas variables independentes sobre unha variable dependente. Aínda que a regresión en media axustada polo método de mínimos cadrados tivo unha gran difusión no século XX, os procedementos de regresión cuantil acadaron un considerable desenvolvemento dende o xurdimiento da Estatística Robusta, que alcanzou a súa máxima expansión nos anos 80. O obxectivo deste traballo é ilustrar as principais vantaxes da regresión cuantil que motivan o interese que suscitan estes modelos no campo da Estatística. Destaca que a regresión cuantil permite obter unha descrición detallada da distribución condicional da variable dependente, non precisa asumir que a distribución do erro é gaussiana e presenta propiedades de robustez. Ilustráronse ditas propiedades tanto dende un punto de vista teórico como práctico grazas aos estudos de simulación e aplicacións a datos reais levados a cabo ao longo do traballo.

**F**erramentas estatísticas para o tratamento de datos de estrelas binarias da misión astrométrica GAIA

**Ángel López Oriona.**

**Titor: César A. Sánchez Sellero.**

A misión Gaia, lanzada pola Axencia Espacial Europea no ano 2013, consiste nun satélite de observación espacial, cuxo obxectivo é trazar un mapa tridimensional da nosa galaxia, a Vía Láctea, cunha precisión nunca antes acadada. Tomamos datos de catálogos estelares obtidos por esta misión, así como de catálogos previos xa validados, como o correspondente a outro satélite anterior, o Hipparcos. Centrámolos no estudo das estrelas dobres, pois crese que constitúen aproximadamente un 70 % das estrelas totais da Vía Láctea. Realizamos unha validación dos datos da misión Gaia relativos a estrelas dobres, para o que se recorreu a contrastes estatísticos de distribucións. Tamén se estudou o problema da estimación da distancia en Astronomía, que entraña unha complexidade singular, e cuxa posible solución pasa por un enfoque probabilístico *bayesiano*.

**M**odelos de regresión con penalizacións

**Sergio Mejuto Vázquez.**

**Titor: Manuel Febrero Bande.**

Os modelos de regresión baséanse no estimador de mínimos cadrados. Pódese ver que é fácil que non sexa capaz de estimar os coeficientes con fiabilidade, no caso de variantes dependentes. Para corrixir este problema, créanse distintos métodos que veremos ao longo deste traballo. A Regresión Ridge e o método LASSO utilizan o estimador de mínimos cadrados cunha penalización que restrinxe os posibles valores que poden tomar os coeficientes e, así, logran controlar os seus valores. Por outra parte, o método LAR dános unha ordenación das variables segundo a súa importancia que nos permite crear modelos de regresión collendo só as variables máis importantes. Con estes métodos conseguimos evitar o problema das variables dependentes.

**P**roblemas de emparellamento

**Estela Vázquez-Monjardín Lorenzo.**

**Titor: Julio González Díaz.**

Un problema de emparellamento é un problema de optimización en redes que busca asociar unha serie de obxectos entre si de maneira que se acade un obxectivo. Dependendo de como estean ordenados estes obxectos, pódese distinguir entre problemas bipartitos e non bipartitos mentres que, dependendo do obxectivo do emparellamento, podemos ter un problema ponderado ou un de máxima cardinalidade. Neste traballo revísanse varios algoritmos para a resolución das distintas versións dos problemas de emparellamento, así como os resultados teóricos necesarios para concluír que atopan solucións óptimas e a velocidade á cal o fan. Ademais, preséntase un algoritmo especialmente deseñado para a versión bipartita de máxima cardinalidade baseado no teorema do camiño aumentador. Aínda que non é moi eficiente, esta técnica de resolución permítenos desenvolver un algoritmo para a versión máis xeral deste problema (a non bipartita) despois dunha serie de sinxelas modificacións.

# Coñecéndonos

## Coñecendo a unha socia da SGAPEIO: Laura Calaza Díaz.

**L**aura Calaza Díaz, é Licenciada en Matemáticas pola Universidade de Santiago de Compostela (USC). Realizou o mestrado en Técnicas Estatísticas e Investigación Operativa e o mestrado de Profesorado de Educación Secundaria Obrigatoria e Bacharelato, Formación Profesional e Ensinanza de Idiomas, ambos os dous pola USC. Tras uns anos traballando como investigadora na USC comezou a súa andaina como profesora de ensino secundario en Galicia no curso 2018/2019.



### Cantos anos levas traballando en Estatística ou IO?

Dende o curso 2014/2015, cando comecei a traballar no grupo de Bioestatística do Departamento de Estatística, Análise Matemática e Optimización da USC.

### Descríbenos o teu traballo

Hai aproximadamente un ano que son profesora de secundaria, na especialidade de Matemáticas, aínda que, neste curso académico, estou exercendo como asesora de formación do profesorado no Centro de Formación e Recursos de Lugo.

### Dende cando es socio da SGAPEIO?

Comecei a ser socia apenas rematei o Mestrado de Técnicas Estatísticas no ano 2015. Pareceume unha moi boa opción para estar en contacto con profesionais da Estatística que traballan en distintos ámbitos e para estar en continua actualización sobre as temáticas que se están a traballar na nosa terra.

### Cal foi o 1º Congreso da SGAPEIO ao que asistiches?

O XII Congreso, que tivo lugar en Lugo no ano 2015.

### O tema do teu próximo proxecto/artigo?

Neste momento non estou adicándolle tantas horas como me gustaría á investigación, dado que o traballo e as actividades extras que sempre acaban aparecendo consomen o tempo máis rápido do que me gustaría. De todos os xeitos, temos pendente de rematar un artigo sobre a análise estatística de datos de dialectoloxía perceptiva empregando modelos mixtos.

### Un curso que che gustaría facer?

Polo que comentei anteriormente...estaría moi ben un curso de xestión do tempo!

### Que libro de Estatística ou I.O. recomendarías?

Un dos grandes referentes, o libro Principios de Inferencia Estatística de Vélez-Ibarrola e García-Pérez.

### Se non foses estatístico, que che gustaría ser?

Unha das miñas paixóns é o deseño. A xestión do espazo e a estética é algo que sempre me resultou atractivo, así que, diría que deseñadora de interiores.

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

Creo que depende do contexto no que a escoite...pensaría en datos, gráficos, información, análise...

### Que é o que máis valoras do labor da SGAPEIO?

Sen dúbida, diría que o seu labor de divulgación da Estatística e a I.O. Estanse organizando continuamente actividades para todos os niveis, desde as que achegan a Estatística aos/ás máis novos/as dun xeito atractivo ata as que permiten actualizar os coñecementos dos/-das máis expertos/as.

### En que cres que deberíamos mellorar?

A Sociedade está a facer un labor de divulgación e de transferencia incansable, iso, sen dúbida. Porén, se tivese que pedir algunha mellora, como docente no Ensino Medio, gustaríame que se realizasen máis actividades ou talleres para acercar ao alumnado o traballo que se está a facer na Universidade, no IGE... que puidesen observar as múltiples aplicacións da Estatística e a súa conexión con outros ámbitos.



# Estatística recreativa

A caixa de pímulas por Antonio Vaamonde Liste

Pilar ten unha caixa con 15 pímulas para o colesterol e, durante todo o mes, ten que tomar media pímula cada noite. Sempre colle ao chou, sen mirar: se sae unha pímula enteira, colle a metade e devolve á caixa a outra metade, e, se sae media pímula (con igual probabilidade que a pímula enteira), tómaa e non devolve nada á caixa.

Cantas medias pímulas é de esperar que queden na caixa o día que colle a derradeira pímula enteira?

A solución estará dispoñible, proximamente, nesta [ligazón](#). Permanecedes atentos!

## Dirección:

Paula Raña Míguez

## Equipo editorial:

M<sup>a</sup> José Ginzo Villamayor, María Gómez Rúa,  
María Martín Vila, Carlos Neira Cortizas,  
Javier Roca Pardiñas,  
Covadonga Rodríguez-Moldes Rey,  
Alejandro Saavedra Nieves,  
César Andrés Sánchez Sellero e  
Javier Tarrío Saavedra,

## Revisión lingüística:

María Miragaya Pereira

## Coa colaboración de:

Ana Belén Buide Carballosa,  
Manuel Antonio Novo Pérez,  
Manuel Oviedo de la Fuente,  
Rosa Crujeiras Casais,  
Laura Calaza Díaz e  
Antonio Vaamonde Liste.

## COLABORA!!!

Informest é unha publicación cuatrimestral da SGAPEIO que podes atopar na web da sociedade, [www.sgapeio.es/](http://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](mailto:boletin@sgapeio.es) onde te informaremos das normas de edición.

**Anímate!!**

## Comunicación coa SGAPEIO:

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

[www.sgapeio.es](http://www.sgapeio.es)

[secretaria@sgapeio.es](mailto:secretaria@sgapeio.es) [boletin@sgapeio.es](mailto:boletin@sgapeio.es)

Twitter @sgapeio

LinkedIn SGAPEIO

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

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