

X Jornadas de usuari@s de MURCIA 2018



X Jornadas de Usuarios de R, Murcia 2018.

Antonio Maurandi López, Aurora González Vidal (Eds.).

Universidad de Murcia.

Digitum, Repositorio Institucional de la Universidad de Murcia, 2020.



© DIGITUM (Biblioteca Universitaria de la Universidad de Murcia), 2020

ISBN: 978-84-09-16357-1

Murcia, 2020.



X Jornadas de Usuarios de R, Murcia 2018

Antonio Maurandi López, Aurora González Vidal



Índice general

1. Las Jornadas de R	5
2. Organización	9
Entidades Organizadoras	9
Comité Organizador	9
Comité Académico	11
Patrocinadores	12
Formato	14
Información de interés	14
Precios y algunos datos	15
3. Programa	17
4. Resúmenes de ponencias	23
Premio a la Mejor Comunicación de Investigadores/as Jóvenes	23
Comunicaciones A1	24
A101. ¿Es posible inferir el comportamiento de un retweet?	24
A102. FactoMineR en tiempos modernos: Un ejemplo	25
A103. Sistema de predicción de alimentos lácteos perecederos	25
A104. Algo de diversión con deep learning: Neuro Style Transfer	25
Comunicaciones A2	27
A201. Pattern of Comorbidities and Associated Risk Factors among Colorectal Cancer Patients in Spain: CoMCoR study	27
A202. Role of R in searching biomarkers for asthma and other respiratory diseases in early childhood	28
A203. Colliders in Applied Biomedical Research: an educational interactive web application	29
A204. Criterios de agrupación de datos en tablas de frecuencias: una validación numérica	30
A205. Herramienta de análisis y minado de workflows en ejercicios de laparoscopia	31
Comunicaciones A3	33
A301. Efecto de las Áreas Marinas Protegidas sobre las poblaciones de peces: aplicación de modelos lineales generalizados mixtos (GLMMs) mediante los paquetes glmer y glmmTMB en R	33
A302. Análítica predictiva en averías de componentes electrónicos	34
A303. Visualización y Monitorización de Servicios con R y Shiny	34
A304. Predicción e interpolación dinámica de los niveles de contaminación en base al tráfico y la dirección del viento	35
Ponencia Invitada I	37
From the package FactoMineR to a project on exploratory multivariate analysis or how to improve the visibility of its package	37
Comunicaciones B	38
B01. rjazz, an R package to control Jazz, the lightweight data processing framework.	38
B02. Time series model selection from a bayessian network practical approach	39
B03. Desarrollo de una aplicación web para el conjuntos de datos de impactron y post-aws	40
B04. Comparación de modelos espacio-temporales de datos de panel con R	41
B05. Anticipar el protocolo de contaminación usando el portal de datos abiertos Madrid	42

B06. Agrupamiento con restricciones para configurar sesiones en conferencias científicas con R + Shiny	42
B07. Text Mining para mejorar nuestro Customer Relationship Management (CRM)	43
Comunicaciones C1	44
C101. RSGIS, una aplicación shiny basada en Sistemas de Información Geográfica	44
C102. Structural Equation Modeling with lavaan to explain the effects of co-creation on consumer behavior	45
C103. deaR. Data Envelopment Analysis con R.	46
C104. Análisis de la influencia de amigos en Facebook con R	46
C105. Applying Data Science to Fraud Prevention and Detection: SALER Analytics	47
Comunicaciones C2	49
C201. Desarrollo de materiales y aplicaciones científicas interactivas en entorno web para la resolución de problemas de optimización	49
C202. Explorando los datos públicos de España: un paquete para navegar la API de datos.gob.es	50
C202. Explorando los datos públicos de España: un paquete para navegar la API de datos.gob.es	51
C203. Bringing Standards to Quality Control with R	52
C204. The efficiency of public lighting service in Spain	54
C204. LAU2boundaries4spain: un paquete de R con contornos municipales para el periodo 2012-2018	54
Ponencia Invitada II	56
Text Mining Using Tidy Data Principles	56
5. Talleres	57
Taller J1: Creación de mapas interactivos con R y Leaflet	57
Taller J2: Evaluación de las medidas de capacidad predictiva	58
Taller J3: Redes Bayesianas con R	58
Taller V1: Text Mining & NLP con R: Desde el análisis teóricos de textos en Esperanto hasta las Aplicaciones Empresariales con shiny	59
Taller V2: Análisis de correspondencias con R	59
6. Pósters	61
Póster 1. Intelligent Climate. A Machine Learning Solution built on SAP Leonardo & R for HVAC Energy and Comfort optimization	61
Póster 2. Evaluación del potencial cartográfico	62
Póster 3. Relación entre el perfil protéico y el intervalo de la muerte en restos óseos humanos	63
Póster 4. Aprendizaje usando las lecciones interactivas	64
Póster 5. Distribuciones tipo fase en memorias RRAM con R	65
Póster 6. Uso de aplicaciones interactivas Shiny	66
Póster 7. Docencia con R en el departamento de Métodos	67
Póster 8. 10 años de Jornadas de usuari@s de R en España	69
Póster 9. Efecto anticoccidial del extracto de ajo en cabras	71
7. Las Jornadas en imágenes	73
8. Repercusión en medios	75

1

Las Jornadas de R

¡Han concluido las décimas Jornadas de R!, las XjuR, como hemos terminado llamándolas. No solo han pasado diez años desde aquellas “míticas” jornadas del 2009 realizadas también en nuestra casa, sino que en estos momentos que escribimos estas líneas van a realizarse las decimoprimeras en Madrid (las XI juR). La sensación cuando hicimos el cierre, con toque musical y entrega de premios, era de júbilo y exaltación de la amistad como en las fiestas adolescentes. No se trató pues de algo habitual, sino de una observación atípica en el universo de cierres de jornadas. Todos vamos a muchos congresos académicos y en general nos despedimos de una forma discreta con un *hasta la vista*. Aquí no ocurre eso, no ha ocurrido en ninguna de las diez que hemos hecho. Las Jornadas de R siguen siendo un encuentro de gente que quiere verse, que quiere compartir, que disfruta con cada charla, cada acto y que se siente partícipe. Quizás el rasgo más que nos gusta señalar de las Jornadas es que *son nuestras* y eso no es cosa baladí. Nos gusta pensar que el espíritu de R está diluido en ellas, R es software libre, es de quien lo usa... como las Jornadas. R lo tomas, lo usas y lo adaptas, lo haces tuyo, así como las Jornadas no son ajenas a los participantes y eso se nota. Se han convertido en una cita anual abierta a todo el mundo, a la que muchos esperamos con ansia y que brillan por ser multidisciplinares, digno reflejo de R.



Además, ya podemos decir que son un clásico del panorama de encuentros dedicados a este mundillo en nuestro país. Creemos que las más *viejas*, aunque también son *jóvenes* igual que sus asistentes, y otros en mentalidad despierta y ágil, pero todos jóvenes.

No hace tanto se cuestionaba la utilidad de unas jornadas exclusivas en torno a R, “*un lenguaje de programación minoritario*”. Ya no es en absoluto así puesto que muchos rankings de lenguajes de programación lo sitúan en ocasiones entre los diez primeros y dada la naturaleza tan específica de R creemos que es todo un logro. Los usuarios y usuarias de R sabemos que no es solo un lenguaje de programación para analizar datos. Hoy en día R es *todo un ecosistema* dentro del cual se puede vivir completamente, donde se puede hacer casi todo de forma casi autónoma y donde además nunca estás sol@. Pero aún más importante, creemos que R representa una forma de pensar que te permite llegar muy lejos. Esperamos que haya otra generación de científic@s, después de las generaciones que se están formando trabajando con R que sean conscientes de que “*R elimina los límites de lo que puedes hacer*”, ya que nunca mejor dicho ahora “*los límites los pones tú*”. Todo esto hace que en los últimos años, R se haya convertido en un estándar, no el único, para sacar partido a los datos que hoy en día son tan y tan abundantes. En el mundo académico es sin lugar a dudas el rey, no se le puede hacer sombra, las posibilidades que ofrece para crear, desarrollar, compartir y recibir feedback son prácticamente inagotables. Y también ocurre lo que siempre ocurre en el mundo del software libre, que en realidad no existe un estándar que lo cubra todo y quizás llegue un día que exista algo mejor y tengamos que migrar... , en absoluto vemos cercano ese día. De momento las Jornadas son exclusivas **no en torno a R** sino de *R* y *SU* entorno que no es poco y es un marco en que cabe todo incluso este conjunto de gentes entusiastas y variopintas.

Este pasado año, en la apertura de las XjuR quisimos hacer un repaso a esta historia que es nuestra. Hace 10 años R éramos

muy pocos los erreros y erreras *sin hache*, estábamos muy aislados e incluso éramos tachados de *rarunos/as*, nos sentíamos algo así como la *aldea de los irreductibles galos*. Pero al juntarnos en Murcia se plantó una semilla que no ha parado de crecer. En aquellas Jornadas del 2009 se escucharon literalmente gritos de “*¡no estamos solos!*” o “*¡no somos tan raros!*”. Era el momento de despegue de R no solo en España sino en el mundo entero y desde entonces ha sido imparable, R está en todas partes. Tuvimos la suerte de celebrar la primera reunión en la Universidad de Murcia, de la mano del profesor José Antonio Palazón que con su espíritu joven y abierto quiso juntar a toda esa gente que no sabía muy bien a que venía, pero que sí sabían que querían venir. Nos conocíamos solo por la lista de correo *r-help.es* que había surgido ese mismo año. Al celebrar las décimas en Murcia quisimos traerlas de vuelta a su *cuna* tras su periplo nacional, con mucha emoción. Y haciendo repaso histórico no podíamos dejar de nombrar al profesor Palazón.

Pensamos que en gran medida todo esto se lo debemos a ese impulso inicial que José Antonio Palazón, *Palax* para los amigos, supo darle a las jornadas allá por el año 2009. La labor de Palax no se quedó allí, sino que a lo largo del tiempo ha creado un grupo de trabajo y de *amig@s*. Nos contagió con su espíritu, *alocado* de compartir, de horizontalidad, de no apuntarse méritos personales sino grupales y de divertirnos con lo que hacíamos. Esta idea ha seguido presente en las Jornadas a lo largo de los años y en sus propias palabras el día de la inauguración resumen el orgullo que todos sentimos al ver las jornadas “*andar solas*, dijo: *el regalo sois vosotros, que esteis aquí, que podamos hacer cosas y que haya futuro*. El éxito de las Jornadas es un logro conjunto de la comunidad, pues era eso lo que se pretendía.



Éste grupo de amigos, cuya semilla plantó Palax, ha ido evolucionando en sí mismo. Primero fuimos el *00Rteam*, un equipo pequeño de fanáticos de los datos que se dedicaba básicamente a evangelizar y difundir R por doquier mediante cursos de doctorado, cursos de estudios propios, talleres *Caldum*¹... pero de una forma esporádica y dependiendo de la disponibilidad de cada uno. En las jornadas de Granada se generó una sinergia muy bonita entre este grupo y un grupo de matemáticas procedente de la Facultad de Economía. De este encuentro ha surgido la Asociación UMUR, *Usuari@s Murcia R*, cuya primera actividad oficial fue su participación tan activa en la organización de las *XjuR*. En UMUR somos amigos, somos compañeros y tenemos una historia. Hoy día organizamos talleres cada dos meses, de forma un poco más ordenada, pero sentimos que nuestro alcance es mayor, a un público más diverso. Somos una asociación nueva pero ya con tablas.

También es destacable la fuerza femenina de R en Murcia. Nos encanta hacer notar que en la junta directiva de UMUR, el ratio mujer:hombre es 7:5, por lo que las mujeres tienen un 70 % de representación en UMUR, rompiendo con el paradigma de que estén infrarrepresentadas en los ámbitos tecnológicos y en los puestos importantes. En esta misma línea, las *XjuR* se adhirieron al manifiesto *No Sin Mujeres* y nos esforzamos por traer una ponente, Julia Silge, y un ponente, Francois Husson, como *keynote speakers*. Por otra parte decir que el esfuerzo fue mínimo, pues hay una gran cantidad de mujeres realizando trabajos destacables en el mundo de R. Estamos orgullosos de mostrar con números que tanto las *XjuR* como en UMUR velamos por sacar adelante actividades de forma inclusiva y con perspectiva de género.

Nos gusta pensar que las jornadas seguirán siendo por mucho tiempo un punto de encuentro y de intercambio de ideas entre personas del mundo empresarial, consultores, analistas... científicos de datos, y del mundo académico (investigadores/as, docentes y estudiantes). En las jornadas se encuentran muchos puntos de vista diferentes dado que venimos de ámbitos muy diversos y nos relajamos al no percibir esa competencia mal entendida. Somos *matemátic@s*, *médic@s*, *biotecnólog@s*, *economistas*, *infomatic@s*, *psicólg@s*, *educadores*, *telecos*, *veterinari@s*, *biólog@s*,... etc. Sabemos que somos diferentes pero que compartimos una afición tremenda por aprender, por saber qué hacen los demás y cómo lo hacen y se aprecia una inmensa generosidad. Desde el principio hemos disfrutado de juntarnos y de compartir. Creemos que la interdisciplinabilidad y la generosidad son las claves de su éxito.

Esperamos que las Jornadas, que ya van por su decimoprimer edición que se celebra en Madrid, sigan siendo ese punto de encuentro extremadamente abierto, muy transversal e intergeneracional y que sigan evolucionando para continuar siendo un evento extraordinario.

Antonio Maurandi y Aurora González

¹Caldum: Conocimiento abierto y libre en la Univ. de Murcia

1. LAS JORNADAS DE R



Figura 1.1: Inauguración de las X Jornadas de Usuari@s de R. Reconocimiento honorífico al Profesor José Antonio Palazón Ferrando.



● INICIO PRESENTACIÓN ASISTE PROGRAMA ORGANIZACIÓN PATROCINIO CONTACTO

X JORNADAS DE USUARIOS DE R

Murcia, 22 y 23 de noviembre de 2018



2

Organización

Entidades Organizadoras



- Comunidad RHispano: La asociación de usuarios de R de España, <http://r-es.org/>
- Facultad de Economía y Empresa, Universidad de Murcia
- UMUR. Asociación de Usuarios de R Murcia, <http://gauss.inf.um.es/umur/>

Comité Organizador

- Alejandro Cisterna García. Universidad de Murcia
- Álvaro Hernández Vicente. Universidad de Murcia
- Ana Belén Marín Valverde. Universidad de Murcia
- Antonio José Perán. Universidad de Murcia
- Antonio Maurandi López (**Coordinador**). Universidad de Murcia, amaurandi@um.es
- Aurora González Vidal. Universidad de Murcia
- Carlos J. Gil Bellosta. RHispano
- Elvira Ferre Jaén. Universidad de Murcia
- Fernando Pérez. Universidad de Murcia
- Francisca Carreño. Universidad de Murcia

- Francisco Javier Ibáñez López. Universidad de Murcia
- Francisco López Martínez. Universidad de Murcia
- Fuensanta Arnaldos. Universidad de Murcia
- Fuensanta Monroy. Universidad de Murcia
- Isabel Parra Frutos. Universidad de Murcia
- José Antonio Sánchez Martí. Universidad de Murcia
- Juan José Pérez Castejón. Universidad de Murcia
- Laura Antón Sánchez. Universidad de Murcia
- Laura del Río Alonso. Universidad de Murcia
- Lourdes Molera. Universidad de Murcia
- María Teresa Díaz. Universidad de Murcia
- Rafael López Serrano. Cajamar
- Ursula Faura. Universidad de Murcia
- Virgilio Gómez Rubio. Universidad de Castilla la Mancha



Figura 2.1: Regalicos!

2. ORGANIZACIÓN

Comité Académico

- Antonio Maurandi López
- Aurora González Vidal
- Emilio Torres Manzanera
- Francisca Carreño
- Francisco Alonso Sarria
- José Antonio Palazón Ferrando
- José Luis Cañadas Reche (**Coordinador**), canadasreche@gmail.com
- Juan José Pérez Castejón
- Laura del Río Alonso
- Leonardo Hansa
- Manuel Madrigal Torres
- Marcos Marvá Ruiz
- Miguel Ángel Rodríguez Muñíos
- Pedro Concejero
- Pilar Olivares Carrillo



Figura 2.2: Parte del Comité Organizador y algunos asiduos de las jornadas.

Patrocinadores

Patrocinadores **Platinum**:

- Grupo Cooperativo CAJAMAR
- RConsortium
- RStudio

Patrocinadores **Oro**:

- Innova-tsn
- Máxima Formación
- UniData
- Cátedra Prosur de biotecnología de alimentos, Universidad de Murcia
- Balneario de Archena
- Springer
- Orange
- Vicerrectorado de Estudiantes, Universidad de Murcia
- Facultad de Economía y Empresa, Universidad de Murcia
- Facultad de Educación, Universidad de Murcia
- Facultad de Matemáticas, Universidad de Murcia
- Facultad de Informática, Universidad de Murcia
- ACTI: Área Científica y Técnica de Investigación, Universidad de Murcia
- El grupo de Innovación docente de la Universidad de Murcia: “GID Estadística en Ciencias Sociales”
- El grupo de Innovación docente de la Universidad de Murcia: “GIDMUR. Grupo de innovación docente multidisciplinar de la Universidad de Murcia”



2. ORGANIZACIÓN



Figura 2.3: Tarjetas identificadoras de asistentes.

Formato

Las jornadas son el reflejo de la importancia creciente de R. Están abiertas a usuarios y entusiastas de R independientemente de su filiación o área de interés. Buscan, además, fomentar las relaciones sociales entre los usuarios de R.

Objetivos:

- Proporcionar un punto de encuentro a los usuarios de R
- Fomentar la colaboración en un ambiente multidisciplinar
- Divulgar el conocimiento del lenguaje R y sus posibilidades
- Promover el uso de R

Formato

- Ponencias invitadas
- Presentaciones de usuarios de R que reflejen el amplio rango en que se está utilizando R para analizar datos
- Sesiones de pósters
- Tutoriales que profundizan en aspectos interesantes y menos conocidos de R

Información de interés

- **Facultad de Economía y Empresa:** La recogida de acreditaciones y recepción será el vestíbulo del sótano de la Facultad de Economía y Empresa. Edificio 2, Campus de Espinardo. Ver mapa aquí: http://www.accessia.es/#Facultad_de_Econom%25EDa_y_Empresa.
- **Aparcamiento:** Recomendamos el aparcamiento que hay junto a la pista de atletismo.
- **Paradas de tranvía:** desde la parada Servicios de investigación atravesando el aparcamiento subes a la FEE.
- **Comidas:** Comedor de la Facultad de Economía y Empresa, vestíbulo sótano. Están incluidas con la inscripción, recoge tus vales en la acreditación.
- **Cena social:** Restaurante Madre de Dios: C/ Vara de Rey, 11, 30001 Murcia apúntate a la cena aquí: <https://encuestas.um.es/encuestas/xjurcena.cc>
- **Concierto Jueves noche:** Sala Revolver, C/ Victorio, 36, 30003 Murcia, <https://goo.gl/maps/fDuVcJNcXDn>



Figura 2.4: Plano del Campus de Espinardo con indicaciones para llegar al Congreso.

2. ORGANIZACIÓN



- @xjurum
- amaurandi@um.es
- <http://r-es.org/XjuR/>
- http://gauss.inf.um.es/umur/actividades.html#2_actividades_del_2018

Precios y algunos datos

El precio de las jornadas osciló entre 35€ y 45€. La mayor parte de los asistentes pagaron inscripción reducida antes del 1 de noviembre.

- Número de asistentes: 144
- Número de ponencias: 32
- Número de talleres: 5
- Número de posters: 9
- Número de premios: 1
- Número de ponencias invitadas: 2

Precios		
	[11 de Abril , 1 de Noviembre)	[1 de Noviembre , 18 de Noviembre]
Reducida	35€	45€
General	45€	55€
Cena	25€. Reserva aquí . Lugar Restaurante Madre de Dios en Murcia .	

Figura 2.5: Precios de las jornadas anunciados en la web <http://r-es.org/XjuR/>.

3

Programa

Programa Jueves 22/nov/18		
9:00-9:30	Entrega de documentación y acreditación	Vestíbulo sótano
9:30-10:20	Inauguración Oficial de las Jornadas	Salón de Actos
10:20-11:20	Comunicaciones paralelas A1, A2 y A3	Sala de Grados y Aulas F0.17 y F0.18
11:20-12:00	Café de bienvenida	Vestíbulo sótano
12:00-13:00 Salón de Actos	Ponencia invitada I <i>From the package FactoMineR to a project on exploratory multivariate analysis or how to improve the visibility of its package</i> François Husson	
13:00-14:30	Comunicaciones B	Salón de Actos
14:30-16:30	Comida	Comedor, vestíbulo sótano
16:30-19:00	Talleres paralelos J1, J2 y J3	
19:00-20:00	Asamblea de R-Hispano	Sala de Grados
21:00	Cena social	Rest. Madre de Dios, Murcia

Programa		
Viernes 23/nov/18		
9:30-10:30	Sesión de posters y café	Vestíbulo sótano
10:30-13:00	Talleres Parelos V1 y V2	
13:00-14:00	Comunicaciones paralelas C1 y C2	Sala de Grados (FEE)y Aula de Centro I (F.Educación)
14:00-16:00	Comida	Comedor, vestíbulo sótano
16:00-17:00	Ponencia invitada II <i>Text Mining Using Tidy Data Principles</i>	
Salón de Actos	Julia Silge (streaming)	
17:00-17:30	Café despedida	Vestíbulo sótano
17:30-18:00	Entrega de premios	Salón Actos
18:00-19:00	Cierre de Jornadas con sorpresa musical	Salón Actos

10:20 a 11:20	Comunicaciones A1, Salón de Grados	Autorías	Entidad
Comunicaciones A1 Coord. Aurora González	¿Es posible inferir el comportamiento de un retweet?	Rafael López Serrano	Univ. de Murcia
	FactoMineR en tiempos modernos: Un ejemplo	José Luis Cañadas Reche	ORANGE España
	Sistema de predicción de alimentos lácteos perecederos	Emilio Torres Manzanera, Carlos Javier Gil Bellosta, Luz Diaz Frías, Susana Montes Rodríguez	Univ. de Oviedo y Circiter
	Algo de diversión con deep learning: Neuro Style Transfer	Pedro Concejero Cerezo	Univ. Complutense de Madrid

10:20 a 11:20	Comunicaciones A2, Aula F0.18	Autorías	Entidad
Comunicaciones A2 Coord. Francisca Carreño	Pattern of Comorbidities and Associated Risk Factors among Colorectal Cancer Patients in Spain: CoMCoR study	Miguel Angel Luque-Fernandez, Daniel Redondo Sánchez, Miguel Rodríguez Barranco, M ^a Carmen Carmona-García, Rafael Marcos Gragera, María José Sánchez Pérez	Univ. de Granada, ibs.Granada y CIBERESP
	Role of R in searching biomarkers for asthma and other respiratory diseases in early childhood	Rosa Alba Sola Martínez, José María Pastor Hernández, Esther Cantero-Cano, Manuel Cánovas Díaz y Teresa de Diego Puente	Univ. de Murcia y IMIB-Arrixaca, Murcia
	Criterios de agrupación de datos en tablas de frecuencias: una validación numérica	Francisco G. Morillas Jurado, Josep Lledó Benito y José M. Pavía	Univ. de Valencia y Univ. de Alcalá de Henares
	Herramienta de análisis y minado de workflows en ejercicios de laparoscopia	David Nieves, Cèsar Ferri, José Hernández-Orallo, Carlos Monserrat y María José Ramírez-Quintana	Univ. Politécnica de Valencia

3. PROGRAMA

10:20 a 11:20			
Comunicaciones A3, Aula F0.17		Autorias	Entidad
Comunicaciones A1 Coord. Francisco Ibáñez López	Efecto de las Áreas Marinas Protegidas sobre las poblaciones de peces: aplicación de modelos lineales generalizados mixtos (GLMMs) mediante los paquetes glmer y glmmTMB en R	Irene Rojo Moreno	Dept de Ecología e Hidrología. Univ. de Murcia
	Análítica predictiva en averías de componentes electrónicos	Aleix Cartró Pérez, Jaume Puigbò Sanvisens	Innova-tsn
	Visualización y Monitorización de Servicios con R y Shiny	Jorge Pradas Moscardó	Edicom Capital, S.L.
	Predicción e interpolación dinámica de los niveles de contaminación en base al tráfico y la dirección del viento	Lidia Contreras Ochando y Cèsar Ferri Ramírez	DSIC, Univ. Politécnica de Valencia
13:00 a 14:30			
Comunicaciones B, Salón de Actos		Autorias	Entidad
Comunicaciones B Coord. José Luis Cañadas Reche	rjazz, an R package to control Jazz, the lightweight data processing framework	Santiago Basaldúa	BBVA Data & Analytics
	Time series model selection from a bayessian network practical approach	Ana Arias Botey, Gabriel Valverde Castilla	CUNEF
	Desarrollo de una aplicación web para el conjuntos de datos de impactron y Post-AWS	Andrés Bishop y Enric Aguilar	Centre for Climate Change, Univ. Rovira i Virgili
	Comparación de modelos espacio-temporales de datos de panel con R	Patricia Carracedo y Ana Debón	Univ. Internacional de Valencia y Univ. Politécnica de Valencia
	Anticipar el protocolo de contaminación usando el portal de datos abiertos Madrid	Marta Lopez Ron, María Neira Bustamante, Fernando Sebastián Huerta	INNOVA-TSN
	Agrupamiento con restricciones para configurar sesiones en conferencias científicas con R + Shiny	Cèsar Ferri, Diego Fernando Vallejo Huanga y Paulina Adriana Morillo Alcívar	UPV, Univ. San Francisco de Quito y Univ. Politecnica Salesiana
	Text Mining para mejorar nuestro Customer Relationship Management (CRM)	Carlos Vecina Tebar	Gnarum Tecnología y Energía

13:00 a 14:00	Comunicaciones C1, Salón de Grados	Autorías	Entidad
Comunicaciones C1 Coord. Juan José Pérez Castejón	RSGIS, una aplicación shiny basada en Sistemas de Información Geográfica	José Antonio Sánchez Martí	Univ. de Murcia
	Structural Equation Modeling with lavaan to explain the effects of co-creation on consumer behavior	Rocío Alarcón-López, Inés López-López and Salvador Ruiz-de-Maya	Univ. de Murcia
	deaR. Data Envelopment Analysis con R	Vicente Coll Serrano, Rafael Benítez Suárez y Vicente J. Bolós Lacave	Univ. de Valencia
	Análisis de la influencia de amigos en Facebook con R	Manuela López, Mariola Palazón, María Sicilia	Univ. de Murcia
	Applying Data Science to Fraud Prevention and Detection: SALER Analytics	Fernando Martínez-Plumed, Juan Carlos Casamayor, Cèsar Ferri, Jon Ander Gómez y Eduardo Vendrell	Univ. Politécnica de Valencia

13:00 a 14:00	Comunicaciones C2, Fac Educación, Aula de Centro I	Autorías	Entidad
Comunicaciones C2 P. Olivares y J.A. Palazón	Desarrollo de materiales y aplicaciones científicas interactivas en entorno web para la resolución de problemas de optimización	Manuel Muñoz Márquez	Univ. de Cádiz
	Colliders in Applied Biomedical Research: an educational interactive web application	Miguel Angel Luque-Fernandez, Daniel Redondo Sánchez, Michael Schomaker, Maria José Sánchez Pérez, Anand Vaidya y Mireille E. Schnitzer	Univ. de Granada, ibr.Granada y CIBERESP
	Explorando los datos públicos de España: un paquete para navegar la API de datos.gob.es	Jorge Cimentada, Jorge Lopez	Univ. Pompeu Fabra, Kernel Analytics
	Bringing Standards to Quality Control with R	Manuel Alfaro García, Emilio López Cano	Univ. de Castilla-La Mancha
	The efficiency of public lighting service in Spain	Pedro-José Martínez Córdoba; María-Dolores Guillamón López; Bernardino Benito López; Úrsula Faura Martínez	Univ. de Murcia
	LAU2boundaries4spain: un paquete de R con contornos municipales para el periodo 2002-2018	Pedro J. Pérez, Francisco Goerlich	Univ. de Valencia e IVIE

3. PROGRAMA

Jueves y viernes	Talleres	Ponente	Aula
J1	Creación de mapas interactivos con R y Leaflet	Antonio Sánchez Chinchón (Telefónica)	Aula Canario (5ª planta FEE)
J2	Evaluación de las medidas de capacidad predictiva obtenidas con técnicas de remuestreo mediante un test de significación	Jesús Herranz (GEICAM)	Aula F0.18
J3	Redes Bayesianas con R	Bojan Mihaljević (Univ. Politécnica de Madrid)	Aula Cigüeña (5ª Planta FEE)
V1	Text Mining & NLP con R: Desde el análisis teóricos de textos en Esperanto hasta las Aplicaciones Empresariales con shiny	Francisco Jesús Rodríguez Aragón (Teradata-Think Big Analytics)	Aula Cigüeña (5ª Planta FEE)
V2	Análisis de correspondencias con R: aplicación a (micro) datos de encuestas	Emilio López Cano (Univ. de Castilla-La Mancha)	Aula F0.18

4

Resúmenes de ponencias

Las ponencias se distribuyeron en seis salones. El primer día hubo tres salones paralelos (A1, A2 y A3), un salón único (B) y una ponencia invitada a cargo de François Husson. El segundo día dos salones paralelos (C1 y C2) y una ponencia invitada a cargo de Julia Silge (streaming). Todas ellas excepto el salón C2 que se llevó a cabo en la Facultad de Educación se realizaron en la Facultad de Economía y Empresa de la Universidad de Murcia.

Premio a la Mejor Comunicación de Investigadores/as Jóvenes

Se otorgaron tres premios a dos ponencias de entre todas las ponencias de las jornadas: premio a “*Mejor Comunicación de Investigadores/as Jóvenes*”. Los premios fueron financiados por el Grupo Cooperativo CAJAMAR.

- A **Ana Arias Botey** por la ponencia titulada `Time series model selection from a bayessian network practical approach`, con reconocida autoría de Ana Arias Botey y Gabriel Valverde Castilla.
- A **Jorge Cimentada** por la ponencia titulada `Explorando los datos públicos de España: un paquete para navegar la API de datos.gob.es`, con reconocida autoría de Jorge Cimentada y Jorge López.
- A **Jorge López** por la ponencia titulada `Explorando los datos públicos de España: un paquete para navegar la API de datos.gob.es`, con reconocida autoría de Jorge Cimentada y Jorge López.

Comité premio mejor ponencia: Pedro Concejero, José Luis Cañadas, Lourdes Molera, Antonio Maurandi, Francisco Ibañez, Ursula Faura.



Figura 4.1: Entrega de los premios a las Mejores Comunicaciones de Investigadores/as Jóvenes.

Comunicaciones A1

Cuadro 4.1: **Ponencias Salón A1** (id es un enlace a la presentación).

id	Titulo	Autorías	Entidad
A101	¿Es posible inferir el comportamiento de un retweet?	Rafael López Serrano	Univ. de Murcia
A102	FactoMineR en tiempos modernos: Un ejemplo	José Luis Cañadas Reche	ORANGE España
A103	Sistema de predicción de alimentos lácteos perecederos	Emilio Torres Manzanera, Carlos Javier Gil Bellosta, Luz Díaz Frías, Susana Montes Rodríguez	Univ. de Oviedo y Circiter
A104	Algo de diversión con deep learning: Neuro Style Transfer	Pedro Concejero Cerezo Univ.	Complutense de Madrid

A101. ¿Es posible inferir el comportamiento de un retweet?

Rafael López Serrano^a

^aUniversidad de Murcia, rafael.lopez@um.es

La distribución de frecuencias de los retweets es muy similar entre tweets. No obstante, si fijamos nuestra atención en la cola de dicha distribución, es posible inferir información útil como el tiempo que durará el interés del tweet, el número de personas a las que les llegó, las que mostraron interés frente a las que no, entre otros. Este trabajo se centra en el estudio de la cola de la distribución de frecuencias para determinar qué información es extrapolable a la hora de responder a preguntas como las expuestas anteriormente. ¿Cuáles son las dudas que nos planteamos? En esta ponencia se presentan algunas de las más importantes a las que los investigadores han intentado dar respuesta en los últimos años. Y ¿qué resultados han obtenido? Todo apunta a que las características que hacen que un tweet se convierta en viral pertenecen mucho más al azar que a las propias de mensaje. Ni siquiera es posible saber a los usuarios a los que llegará o los días que estará activo ¿Qué podemos hacer? Los estudios presentados hasta el momento ponen de manifiesto que las características propias del mensaje o la red social no son suficientes para inferir sobre las preguntas planteadas, y que el proceso tiene mucho más de azar que de predictivo debido a la variabilidad que presentan los datos. Parece que aunque no es posible responder a las preguntas planteadas anteriormente, si se puede minimizar el error bajo ciertas suposiciones concretas. En este estudio focalizamos los tweets analizados concretando parámetros sobre su frecuencia y temática. Usando un modelo aditivo, es posible descomponer la serie en su parte estacional, su tendencia y la componente aleatoria de la misma aplicando un modelo de descomposición de series temporales. De esta forma podemos centrar la atención en su tendencia, ya que es la que realmente contiene la información interesante sobre el alcance o duración de los retweets en Twitter, en definitiva sobre su comportamiento. Se observa que la forma de la serie que se presenta es muy característica para la mayor parte de las series de retweets estudiados y tratamos de analizar los resultados definiendo parámetros sobre la misma. En esta ponencia se muestran algunos de los parámetros incluidos en el estudio, como son el número máximo de tweets, el punto de corte de dicho máximo, el valor que supone el Q3 (tercer cuartil) como valor significativo del decremento de la distribución de frecuencias, el número de mensajes iniciales, el alcance en tiempo del tweet, la fase de incremento y decremento de la distribución de frecuencias, entre otros parámetros propios de la distribución de cada retweet. Añadido a estos se calculan valores globales como el número total de días, la fecha inicial y final, la media de retweet por día, como parámetros importantes globales que podrían influir en el modelo de predicción. Todo esto hace que se pueda modelizar el alcance, o algunos parámetros similares de la serie de retweet, obteniendo resultados fiables.

Paquetes de R que emplea: rtweet

Referencias/Bibliografía

- Birkin, Mark and Malleon, Nicholas (2012) Investigating the Behaviour of Twitter Users to Construct an Individual-Level Model of Metropolitan Dynamics.

- Tan, Chenhao & Lee, Lillian & Pang, Bo. (2014). The effect of wording on message propagation: Topic- and author-controlled natural experiments on Twitter. 52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference. 1. 10.3115/v1/P14-1017.
- Quanzhi Li, Xiaomo Liu, Rui Fang, Armineh Nourbakhsh, Sameena Shah (2016). User Behaviors in Newsworthy Rumors: A Case Study of Twitter. AAAI Publications, Tenth International AAAI Conference on Web and Social Media.
- Maximilian Jenders, Gjergji Kasneci, and Felix Naumann. (2013). Analyzing and predicting viral tweets. In Proceedings of the 22nd International Conference on World Wide Web (WWW '13 Companion). ACM, New York, NY, USA, 657-664. DOI: <https://doi.org/10.1145/2487788.2488017>.

A102. FactoMineR en tiempos modernos: Un ejemplo

José Luis Cañadas Reche^a

^aORANGE España, canadasreche@gmail.com

Con el advenimiento de métodos como los word embedding o tecnologías como spark hemos dejado un poco de lado técnicas clásicas como el análisis de correspondencias simple y múltiple, que en esencia persiguen el mismo objetivo que las más actuales basadas en deep learning, a saber, proyectar en un subespacio “óptimo” relaciones entre categorías de variables. En esta ponencia contaré con un ejemplo muy sencillo como se pueden seguir utilizando estas técnicas dentro del mundo del big data

Paquetes de R que emplea: sparklyr, FactoMineR, factoextra, factoinvestigate, cdata, dplyr

A103. Sistema de predicción de alimentos lácteos percederos

Emilio Torres Manzanera^a Carlos Javier Gil Bellosta^b Luz Diaz Frías^c Susana Montes Rodríguez^d

^aUniversidad de Oviedo, torres@uniovi.es

^bCirciter, cgb@circiter.es

^cCirciter, luzfrias@circiter.es

^dUniversidad de Oviedo, montes@uniovi.es

Reny Picot es una empresa española dedicada a la comercialización de productos lácteos, tales como leche, queso o mantequilla. Como política de empresa atiende y despacha de forma inmediata los pedidos de sus clientes, por lo que necesita disponer de mercancía suficiente en el almacén. Al ser productos perecederos, ha de buscar un equilibrio entre lo que produce y lo que realmente vende, con el fin de reducir los costes de rotura. Así, se ha implementado un sistema de predicción y almacenamiento, integrado con las plataformas ofimáticas de la empresa, que determina la previsión de ventas semanales y ofrece una estimación de producción que minimiza el coste de rotura y el volumen de mercancía que se deshecha por caducidad.

A104. Algo de diversión con deep learning: Neuro Style Transfer

Pedro Concejero Cerezo^a

^aUniversidad Complutense de Madrid, pedro.concejero@ucm.es

Esta ponencia pretende introducir un área relativamente novedosa dentro del campo de Machine Learning y que resulta posible por los recientes avances en “deep learning” y “transfer learning”, conceptos ambos que se introducirán y discutirán adecuadamente. Se trata de la posibilidad de crear, o más bien re-crear imágenes a partir de contenido trasladando el estilo de otra imagen. Se basa

en el popular procedimiento de Gatys et al. (2015) quien aplica con bastante fortuna los modelos convolucionales (VGG-19) de Simonyan y Zisserman (2014) para trasladar el estilo por ejemplo de un Van Gogh a una imagen. En la ponencia presentaremos resultados de una aplicación similar para intentar replicar las variaciones sobre el retrato de Inocencio X de Velázquez que propuso Francis Bacon, y muchas otras pruebas con una variedad de imágenes. Todo ello se realiza con la librería keras disponible para R (<https://keras.rstudio.com/>) que ha popularizado enormemente el entrenamiento de modelos de deep learning en la comunidad R. Como parte de la ponencia se plantearán también las posibilidades que tenemos para realizar este tipo de modelos de una forma eficiente, puesto que la complejidad computacional de los algoritmos puede requerir el uso de GPU's.

Paquetes de R que emplea: keras

Referencias/Bibliografía

- A. Z. Karen Simonyan, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv (1409).
- M. B. Leon A. Gatys, Alexander S. Ecker, A neural algorithm of artistic style, arXiv preprint arXiv (1508).

Comunicaciones A2

Cuadro 4.2: **Ponencias Salón A2** (id es un enlace a la presentación).

id	Título	Autorías	Entidad
A201	Pattern of Comorbidities and Associated Risk Factors among Colorectal Cancer Patients in Spain: CoMCoR study	Miguel Ángel Luque-Fernández, Daniel Redondo Sánchez, Miguel Rodríguez Barranco, Ma Carmen Carmona-García, Rafael Marcos Gragera, María José Sánchez Pérez	Univ. de Granada, ibr.Granada y CIBERESP
A202	Role of R in searching biomarkers for asthma and other respiratory diseases in early childhood	Rosa Alba Sola Martínez, José María Pastor Hernández, Esther Cantero-Cano, Manuel Cánovas Díaz y Tersa de Diego Puente	Univ. de Murcia y IMIB-Arrixaca, Murcia
A203	Colliders in Applied Biomedical Research: an educational interactive web application	Miguel Ángel Luque-Fernández, Daniel Redondo Sánchez, Michael Schomaker, Maria José Sánchez Pérez, Anand Vaidya y Mireille E. Schnitzer	Univ. de Granada, ibr.Granada y CIBERESP
A204	Herramienta de análisis y minado de workflows en ejercicios de laparoscopia	David Nieves, Cèsar Ferri, José Hernández-Orallo, Carlos Monserrat y María José Ramírez-Quintana	Univ. Politécnica de Valencia

A201. Pattern of Comorbidities and Associated Risk Factors among Colorectal Cancer Patients in Spain: CoMCoR study

Miguel Angel Luque-Fernandez^a, Daniel Redondo Sánchez^b, Miguel Rodríguez Barranco^c, M^a Carmen Carmona-García^d, Rafael Marcos Gragera^e, María José Sánchez Pérez^f

^aNon-Communicable Disease and Cancer Epidemiology Group, Biomedical Research Institute of Granada, University of Granada

^bNon-Communicable Disease and Cancer Epidemiology Group, Biomedical Research Institute of Granada, University of Granada

^cBiomedical Network Research Centers of Epidemiology and Public Health (CIBERESP), Spain

^dCatalan Institute of Oncology, Barcelona, Spain

^eCatalan Institute of Oncology, Barcelona, Spain

^fNon-Communicable Disease and Cancer Epidemiology Group, Biomedical Research Institute of Granada, University of Granada

Background: Cancer accounted for 8.8 million deaths globally in 2015 and was the second most common cause of death in the world. Colorectal cancer (CRC) is the second most frequently diagnosed cancer in Spain for both men and women with over 41,441 new cases in 2015. Besides clinical and pathological characteristics of the tumor, also general health status and comorbidities can influence cancer treatment and outcomes. Comorbidity and multimorbidity are increasingly seen as a problem of the elderly. A number of studies have been performed analysing the influence of comorbidity on cancer outcomes, but little evidence is available regarding the frequency and distribution of comorbidities at a population level among CRC patients in Spain. Methods: We developed a population-based high-resolution cohort study, including all CRC incident cases diagnosed in 2011 (n= 1,061). Data were drawn from two population cancer registries and patient's medical records. We defined comorbidity as the existence of a long-term health condition or disorder in the presence of cancer, whereas multimorbidity refers to the existence of two or more comorbid conditions. In order to reduce the high dimensionality of the data, we used advanced visualization tools helping to visualize and interpret the data: radar-plots, heatmaps and forest plots for the ten most common comorbidities. We described the frequency and distribution of comorbidities and multimorbidity by patient, tumor and healthcare factors using radar-plots and heatmaps. Then, we used generalized linear models

to characterize the risk factors associated with the presence of the most prevalent comorbidities plus dementia and multimorbidity. We used forest plots to display the results. Finally, we developed an interactive web application <http://watzilei.com/shiny/CoMCoR/> and a GitHub repository <https://github.com/migariane/CoMCoR> for reproducibility and allowing the complete visualization of the study results. Results: The most common comorbidities were diabetes (23.6 %), chronic obstructive pulmonary disease (17.2 %) and congestive heart failure (14.5 %). Dementia was the most common comorbidity among older patients (75+ years) showing a higher proportion (30 %) of late cancer diagnosis (stage IV) and hospital emergency admission (33 %). CRC with dementia had nearly three times higher risk of not being offered surgery treatment (RR: 2.8, 95 %CI: 1.6, 5.0). Older (+75 years) obese male and current smoker, late surgery after cancer diagnosis (more than 60 days) and not being offered a surgical treatment were associated with a higher risk of multimorbidity. Patients with multimorbidity (2 or more comorbidities) aged 75+ years showed a higher prevalence of surgery the same day or the day after hospital emergency admission (37 %). Conclusions: CoMCoR found a relevant pattern in the distribution and frequency of comorbidities and multimorbidity among CRC patients in Spain mostly associated with dementia, advanced age and surgical treatment. CRC frequency of late diagnosis (stage IV) among patients with dementia and the high proportion of older patients not being offered surgical treatment are important findings that require policy actions. All the results from CoMCoR are made open source available in a web application which is meant to serve as a scientific tool supporting evidence-based policymaking aiming to improve comorbid CRC patients' outcomes. Funding information: Carlos III Institute of Health, Grant/Award Number: CP17/00206 and the Andalusian Department of Health, Grant Number: PI-0152/2017.

Paquetes de R que emplea: shiny, radarchart, dplyr, reshape2, ggplot2, plotly

Referencias/Bibliografía

- “Visualize This: The Flowing Data Guide to Design, Visualization, and Statistics, by Nathan Yau. Wiley Publishing, 2011.
- Fundamentals of Data Visualization. Claus O. Wilke. O'Reilly Media, 2018.
- The impact of comorbidity on cancer and its treatment. Sarfati, D., B. Koczwara, and C. Jackson. CA Cancer J Clin, 2016.
- Reproducibility, reliability and validity of population-based administrative health data for the assessment of cancer non-related comorbidities. Camille Maringe, Helen Fowler, Bernard Rachet, Miguel Angel Luque-Fernandez. PLOS ONE, 2017.

A202. Role of R in searching biomarkers for asthma and other respiratory diseases in early childhood

Rosa Alba Sola Martínez^a, José María Pastor Hernández^b, Esther Cantero-Cano^c, Manuel Cánovas Díaz^d y Teresa de Diego Puente^e

^aDepartment of Biochemistry and Molecular Biology B and Immunology, University of Murcia, Murcia, Spain

^bDepartment of Biochemistry and Molecular Biology B and Immunology, University of Murcia, Murcia, Spain

^cGroup of Molecular Biology of Systems, Biomedical Research Institute of Murcia (IMIB-Arixaca), Murcia, Spain

^dDepartment of Biochemistry and Molecular Biology B and Immunology, University of Murcia, Murcia, Spain

^eDepartment of Biochemistry and Molecular Biology B and Immunology, University of Murcia, Murcia, Spain

These days, Asthma is one of chronic inflammatory pathologies which has higher prevalence in childhood. For this reason, an early diagnosis would improve the quality of life of these children. Nevertheless, a non-invasive and fitting method for early diagnosis of asthma in infants has not been developed yet. Volatile organic compounds (VOCs) are produced by different inflammatory and metabolic human pathways. Due to different factors as metabolic state or pathologies among others, the composition and the concentration of VOCs varies interpersonally and intrapersonally. Currently, analysis of volatile organic compounds (VOCs) in exhaled breath has been proposed as promising tool in therapeutic monitoring and diagnosis of certain diseases such as asthma. The Project NELA (Nutrition in Early Life and Asthma), whose main objective is studying the role of nutrition and different factors in appearance of asthma and other respiratory diseases in childhood, has recruited a cohort of almost 800 mother-infant pairs. Profiling of VOCs in exhaled breath of three-month-old children and mothers is performed by a coupled system of thermal desorption and gas chromatography-mass spectrometry (TD-GC-MS) by the biochemistry group of NELA project for the first time. Moreover, room air samples, which are used as blank, are also collected and analysed by TD-GC-MS. Mass spectra are compared with spectra from NIST 14 library so as to identify VOCs. Afterwards, DBSCAN clustering and centroid's Euclidean distance which use frequency of appearance, retention time and peak base of VOCs as variables is performed so as to assess that identification is correct. Finally, machine learning algorithms are carried out by package 'class' to reassign a suitable name to VOCs which are identified incorrectly. Nowadays, R is a fundamental tool in biomedicine and biochemistry, due to its capacity of easily managing a large amount of data. In this project,

R is used in pre-processing and biostatistical analyses. Qualitative variables are analysed by Multiple Correspondence Analysis (MCA), followed by Chi-squared test and Fisher's exact test. Quantitative variables are analysed by Multivariate Data Analysis, such as Principal Component Analysis (PCA), logistic regression and Receiver Operating Characteristic (ROC) curve analysis. A preliminary statistical analysis which involved 201 children and 201 mothers of NELA cohort was carried out by PCA of package 'FactoMineR'. First component obtained by PCA accounts for 37.67% of variance, while 1-10 components altogether account for 96.91%. Although mothers and children had the same contribution in first component, children had more weight in second component and mother in fourth component. Thus, it was possible to distinguish between mothers' and children's breath if dimensions 2-4 were selected. Acetone (VOC with the highest contribution in first component) was elevated in mothers, whereas isopropyl alcohol (VOC with the highest contribution in first component) showed higher levels in infants. In fact, isopropyl alcohol, which is present in cleaning products, can be metabolised to acetone by alcohol dehydrogenase. It could indicate different metabolic or exposition pattern between mothers and children. In conclusion, the results show that analysis of exhaled breath is a potential tool in detection of contaminant exposition and diagnosis or monitoring of asthma and other respiratory diseases.

Paquetes de R que emplea: FactoMineR, class

Referencias/Bibliografía

- J. Dallinga, C. Robroeks, J. Van Berkel, E. Moonen, R. Godschalk, Q. Jöbsis, E. Dompeling, E. Wouters, F. Van Schooten, Volatile organic compounds in exhaled breath as a diagnostic tool for asthma in children, *Clinical & Experimental Allergy* 40 (1) (2010) 68–76.
- A. Amann, B. de Lacy Costello, W. Miekisch, J. Schubert, B. Buszewski, J. Pleil, N. Ratcliffe, T. Risby, The human volatilome: volatile organic compounds (voc) in exhaled breath, skin emanations, urine, feces and saliva, *Journal of breath research* 8 (3) (2014) 034001.
- L. Garcia-Marcos, J. Edwards, E. Kennington, P. Aurora, E. Baraldi, S. Carraro, M. Gappa, R. Louis, A. Moreno-Galdo, D. Peroni, et al., Priorities for future research into asthma diagnostic tools: A pan-eu consensus exercise from the european asthma research innovation partnership (earip), *Clinical & Experimental Allergy* 48 (2) (2018) 104–120.
- F. Husson, J. Josse, Multiple correspondence analysis, *Visualization and Verbalization of Data* (2014) 165–184.
- H.-Y. Kim, Statistical notes for clinical researchers: chi-squared test and fisher's exact test, *Restorative dentistry & endodontics* 42 (2) (2017) 152–155.
- I. Jolliffe, Principal component analysis, in: *International encyclopedia of statistical science*, Springer, 2011, pp.1094–1096.
- D. W. Hosmer Jr, S. Lemeshow, R. X. Sturdivant, *Applied logistic regression*, Vol. 398, John Wiley & Sons, 2013.
- K. Hajian-Tilaki, Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation, *Caspian journal of internal medicine* 4 (2) (2013) 627.
- F. Husson, J. Josse, S. Le, J. Mazet, M. F. Husson, Package 'factominer' (2018).
- R. Slaughter, R. Mason, D. Beasley, J. Vale, L. Schep, Isopropanol poisoning, *Clinical toxicology* 52 (5) (2014) 470–478.

A203. Colliders in Applied Biomedical Research: an educational interactive web application

Miguel Angel Luque-Fernandez^a, Daniel Redondo Sánchez^b, Michael Schomaker^c, Maria Jose Sánchez Pérez^d, Anand Vaidya^e y Mireille E. Schnitzer^f

^aBiomedical Research Institute. Non-Communicable and Cancer Epidemiology Group (ibs.Granada), Andalusian School of Public Health, University of Granada, Granada, Spain

^bBiomedical Research Institute. Non-Communicable and Cancer Epidemiology Group (ibs.Granada), Andalusian School of Public Health, University of Granada, Granada, Spain

^cCentre of Infectious Disease Epidemiology and Research. University of Cape Town, Cape Town, South Africa

^dBiomedical Research Institute. Non-Communicable and Cancer Epidemiology Group (ibs.Granada), Andalusian School of Public Health, University of Granada, Granada, Spain

^eBrigham and Women's Hospital. Harvard Medical School, Harvard University, Boston, MA, USA

^fFaculty of Pharmacy, University of Montreal, Montreal, Canada

Classical biomedical research (Epidemiology / Biostatistics) has focused on the control of confounding but it is only recently that applied researchers have started to focus on the bias produced by colliders. A collider for a certain pair of variables (e.g., an outcome Y and an exposure A) is a third variable (C) that is caused by both. In DAGs terminology, a collider is the variable in the middle of an inverted fork (i.e., the variable C in $A \rightarrow C \leftarrow Y$). Controlling for, or conditioning an analysis on a collider (i.e., through stratification or regression) can introduce a spurious association between its causes. This potentially explains many paradoxical findings in the medical literature, where established risk factors for a particular outcome appear protective. We used an example from non-communicable disease epidemiology to contextualize and explain the effect of conditioning on a collider. We generated a dataset with 1,000 observations and ran Monte-Carlo simulations to estimate the effect of 24-hour dietary sodium intake on systolic blood pressure, controlling for age, which acts as a confounder, and 24-hour urinary protein excretion, which acts as a collider. We illustrate how adding a collider to a regression model introduces bias. Thus, to prevent paradoxical associations, applied researchers estimating causal effects should be wary of conditioning on colliders. We provide R-code in easy-to-read boxes throughout the manuscript and a GitHub repository (<https://github.com/migariane/ColliderApp>) for the reader to reproduce our example. We also provide an educational web application allowing real-time interaction to visualize the paradoxical effect of conditioning on a collider <http://watzilei.com/shiny/collider/>. We investigated a situation where adding a certain type of variable to a linear regression model, called a ζ collider ζ , led to bias with respect to the regression coefficient estimates while still improving the model fit. DAGs are based on subject matter knowledge and are vital for identifying colliders. Determining if a variable is a collider involves critical thinking about the true unobserved data generation process and the relationship between the variables for a given scenario. Then, the decision whether to include or exclude the variable in a regression model using observational data in biomedical research is based on whether the purpose of the study is prediction or explanation/causation. Under the structures we investigated here, adding a collider to a regression model is not advised when one is interested in the estimation of causal effects, as this may open a back-door path. However, if prediction is the purpose of the model, the inclusion of colliders in the models may be advisable if it reduces the model's prediction error. Most research in Epidemiology and Biostatistics tries to explain how the world works (i.e., it is causal), thus to prevent paradoxical associations, applied researchers estimating causal effects should be aware of such variables. Funding information: Carlos III Institute of Health, Grant/Award Number: CP17/00206 and the Andalusian Department of Health, Grant Number: PI-0152/2017.

Paquetes de R que emplea: shiny, ggplot2, dplyr, visreg

Referencias/Bibliografía

- Pearl, J. Causality: models, reasoning, and inference (2nd ed.). Cambridge: Cambridge University Press. 2009.
- Luque-Fernandez MA, Zoega H, Valdimarsdottir U, Williams MA. Deconstructing the smoking-preeclampsia paradox through a counterfactual framework. Eur J Epidemiol. 2016; 31: 613-623.
- Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, et al. Illustrating bias due to conditioning on a collider. Int J Epidemiol. 2009;39:417-420

A204. Criterios de agrupación de datos en tablas de frecuencias: una validación numérica

Francisco G. Morillas Jurado^a, Josep Lledó Benito^b, y José M. Pavía^c

^aDpto. Economía Aplicada de la Universitat de València, Francisco.Morillas@uv.es

^bDpto. Economía y Dirección de Empresas de la Universidad de Alcalá de Henares

^cDpto. Economía Aplicada de la Universitat de València

En la descripción de un fenómeno es habitual resumir los datos obtenidos de maneras diferentes con la finalidad de transformar estos datos en información que sea útil directamente para la toma de decisiones, o como input para la aplicación de metodologías concretas. En este sentido la información puede resumirse en forma de tabla numérica, presentarse de forma gráfica u obteniendo algunas medidas o indicadores de la misma (en este conjunto se incluyen las medidas de posición, dispersión...). En particular, este trabajo se enmarca en la manera de realizar un resumen en forma de tabla de frecuencias de los datos disponibles en cada caso. Así, basado en la discusión realizada con los profesores R. Herrerías Pleguezuelo- J.M. Herrerías Velasco, en el congreso de Econometría y Métodos Cuantitativos IMEQ2015, este trabajo expone algunas reglas clásicas en la elección del número de categorías/intervalos a

considerar para la elaboración de una tabla de frecuencias, y analiza la validez de las mismas para diferentes distribuciones clásicas de probabilidad mediante la utilización de la simulación numérica, considerando, por ejemplo, aspectos como el tamaño de la muestra disponible, ciertas características de forma de la distribución testada (simetría, curtosis. . .); así como la presencia/ausencia de datos extremos/anómalos. Todo ello en el marco de la asignatura de Estadística I que se imparte en la Facultat de Economia de la Universitat de València en grados como ADE, Economía, FyC. . .

A205. Herramienta de análisis y minado de workflows en ejercicios de laparoscopia

David Nieves^a, Cèsar Ferri^b, José Hernández-Orallo^c, Carlos Monserrat^d y María José Ramírez-Quintana^e

^aDepartamento de Sistemas Informáticos y Computación.Universitat Politècnica de València, Valencia (España), daniecor@dsic.upv.es

^bDepartamento de Sistemas Informáticos y Computación.Universitat Politècnica de València, Valencia (España), cferri@dsic.upv.es

^cDepartamento de Sistemas Informáticos y Computación.Universitat Politècnica de València, Valencia (España), jorallo@dsic.upv.es

^dDepartamento de Sistemas Informáticos y Computación.Universitat Politècnica de València, Valencia (España), cmonserr@dsic.upv.es

^eDepartamento de Sistemas Informáticos y Computación.Universitat Politècnica de València, Valencia (España), mramirez@dsic.upv.es

Los avances tecnológicos en aprendizaje automático y visión por ordenador han brindado la oportunidad de desarrollar sistemas de supervisión automática que puedan ayudar a los cirujanos noveles a mejorar sus destrezas mediante la detección y corrección temprana de errores durante sus primeros pasos en la cirugía laparoscópica . En muchos casos, los motivos que se encuentra detrás de una mala intervención quirúrgica se hallan en una adquisición deficiente de estas destrezas, lo que puede acarrear postoperatorios más largos o, en los casos más extremos, la muerte del paciente . Antes de que el sistema pueda supervisar, y para que esta supervisión sea eficaz, el supervisor automático debe de aprender todas las formas correctas en las que la rutina de cirugía se puede desarrollar. Descartada la programación manual, debido al coste de tiempo y esfuerzo que supone, la opción más eficiente según esas restricciones es que el sistema aprenda partir de ejemplos demostrativos realizados por cirujanos expertos. El coste de grabar estas demostraciones, por parte de los expertos, es exigente para ellos, debido alto nivel de pericia que requieren y el tiempo que demandan. Por tanto, se estima que sean muy pocos los ejemplos con los que el sistema debe de generalizar y aprender las formas correctas (i.e., workflows) en las que un ejercicio de laparoscopia se debe de realizar. Esto implica conocer el conjunto de actividades fundamentales que se requieren, así como determinar el orden o dependencias que puedan existir entre ellas. En definitiva, extraer la esencia de la tarea que subyace dentro de las demostraciones que se proporcionan al sistema.

La dificultad inherente de los ejercicios de laparoscopia provoca que en muchas ocasiones se realicen más actividades quirúrgicas de las estrictamente necesarias . Por otra parte, los cirujanos más experimentados suelen desarrollar una manera propia o “estilo” cuando realizan ciertas rutinas de cirugía (e.g., sutura), lo que puede dificultar la generalización cuando se disponen de muy pocos ejemplos con los que aprender la forma de la tarea . Esto lleva a que la agregación completa de todos los ejemplos, sin ningún tipo de filtrado de información, tienda rápidamente a una representación conceptual de la tarea, o *workflow*, que no proporcione ninguna abstracción significativa, y que sea cuasi imposible de interpretar visualmente. En este trabajo, por tanto, presentamos una herramienta desarrollada en R, que permite la extracción de workflows subyacentes en conjuntos de ejemplos demostrativos de ejercicios de laparoscopia. Esta herramienta esta compuesta por un método novedoso de minado de workflows y una aplicación interactiva con la que poder probarlo con ejemplos reales de laparoscopia.

Resumen más amplio aquí: https://drive.google.com/file/d/1am2MYsukQYkxel5vR7wdQHgt-_daT4hw/view?usp=sharing

Paquetes de R que emplea: “igraph”, “visNetwork”, “plyr”, “dplyr”, “shiny”, “shinydashboard”, “shinycssloaders”, “DT”, “ggplot2”

Referencias/Bibliografía

- Y. Gao, S. S. Vedula, C. E. Reiley, N. Ahmidi, B. Varadarajan, H. C. Lin, L. Tao, L. Zappella, B. Béjar, D. D. Yuh, et al., Jhu-isi gesture and skill assessment working set (jigsaws): A surgical activity dataset for human motion modeling, in: MICCAI Workshop: M2CAI, Vol. 3, 2014, p.3.
- N. Ahmidi, L. Tao, S. Sefati, Y. Gao, C. Lea, B. B. Haro, L. Zappella, S. Khudanpur, R. Vidal, G. D. Hager, A dataset and benchmarks for segmentation and recognition of gestures in robotic surgery, IEEE Transactions on Biomedical Engineering 64 (9) (2017) 2025–2041.

- T. Blum, N. Padoy, H. Feußner, N. Navab, Workflow mining for visualization and analysis of surgeries, *International journal of computer assisted radiology and surgery* 3 (5) (2008) 379–386.
- C. W. Günther, W. M. Van Der Aalst, Fuzzy mining–adaptive process simplification based on multi-perspective metrics, in: *International conference on business process management*, Springer, 2007, pp.328–343.
- S. Yang, M. Zhou, S. Chen, X. Dong, O. Ahmed, R. S. Burd, I. Marsic, Medical workflow modeling using alignment-guided state-splitting hmm, in: *Healthcare Informatics (ICHI), 2017 IEEE International Conference on*, IEEE, 2017, pp.144–153.
- R. Agrawal, D. Gunopulos, F. Leymann, Mining process models from workflow logs, in: *International Conference on Extending Database Technology*, Springer, 1998, pp.467–483.
- E. Galbrun, P. Cellier, N. Tatti, A. Termier, B. Crémilleux, Mining periodic patterns with a mdl criterion, *arXiv preprint arXiv:1807.01706*.
- C. Cao, C. L. MacKenzie, S. Payandeh, Task and motion analyses in endoscopic surgery, in: *Proceedings ASME Dynamic Systems and Control Division*, Citeseer, 1996, pp.583–590.
- D. Nieves, C. Ferri, J. Hernández-Orallo, C. Monserrat, Low-level event detection system for minimally-invasive surgery training, in: *Proceedings of the 4th international Workshop on Sensor-based Activity Recognition and Interaction*, ACM, 2017, p.5.
- C. Monserrat, J. Hernández-Orallo, J. Dolz, M. Rupérez, P. Flach, Knowledge acquisition by abduction for skills monitoring: Application to surgical skills, in: *26th International Conference on Inductive Logic Programming, ILP2016*, 2016.
- J. D. Birkmeyer, J. F. Finks, A. O’reilly, M. Oerline, A. M. Carlin, A. R. Nunn, J. Dimick, M. Banerjee, N. J. Birkmeyer, Surgical skill and complication rates after bariatric surgery.

Comunicaciones A3

Cuadro 4.3: Ponencias Salón A3 (id es un enlace a la presentación).

id	Título	Autorías	Entidad
A301	Efecto de las Áreas Marinas Protegidas sobre las poblaciones de peces: aplicación de modelos lineales generalizados mixtos (GLMMs) mediante los paquetes glmer y glmmTMB en R	Irene Rojo Moreno	Dept de Ecología e Hidrología. Univ. de Murcia
A302	Analítica predictiva en averías de componentes electrónicos	Aleix Cartró Pérez, Jaume Puigbò Sanvisens	Innova-tsnInnova-tsn
A303	Visualización y Monitorización de Servicios con R y Shiny	Jorge Pradas Moscardó	Edicom Capital, S.L.
A304	Predicción e interpolación dinámica de los niveles de contaminación en base al tráfico y la dirección del viento	Lidia Contreras Ochando y César Ferri Ramírez	DSIC, Univ. Politécnica de Valencia

A301. Efecto de las Áreas Marinas Protegidas sobre las poblaciones de peces: aplicación de modelos lineales generalizados mixtos (GLMMs) mediante los paquetes glmer y glmmTMB en R

Irene Rojo Moreno^a

^aDepartamento de Ecología e Hidrología. Universidad de Murcia. ire.rojo@gmail.com

Las Áreas Marinas Protegidas (AMPs) son herramientas efectivas para la gestión de la pesca y la conservación de la biodiversidad. Generalmente consisten en zonas de reserva integral, donde cualquier actividad, incluida la pesca, está prohibida, las cuales son reconocidas como muy beneficiosas a nivel ecológico, y zonas de reserva parcial, donde algunas actividades están permitidas, como la pesca artesanal y el buceo recreativo, para las cuales aún existe cierta controversia sobre su eficacia. El objetivo de este estudio fue evaluar el efecto de la producción primaria y la protección sobre la riqueza, densidad y biomasa de peces mediterráneos. Para ello, se realizaron censos de peces en 5 AMPs mediterráneas localizadas a lo largo de un gradiente de producción primaria en ambos niveles de protección (reservas integral y parcial), así como en zonas adyacentes no protegidas. Con el fin de controlar la variabilidad espacial a escala media, en cada AMP se seleccionaron aleatoriamente tres sitios en la reserva integral, 6 en la reserva parcial y 9 en las zonas no protegidas, y se realizaron 3 transectos aleatorios para el censo de peces en cada uno de los sitios. Para el análisis de los datos se aplicaron modelos lineales generalizados mixtos (GLMMs; Zuur et al., 2009) incluyendo los factores Zona [fijo, 5 niveles (Cabo de Gata, Cabo de Palos, Tabarca, Es Freus y Menorca)], Nivel de protección [fijo, tres niveles (reserva integral, reserva parcial, no protegido)], y la interacción entre ambos factores, y Sitio (factor aleatorio). Para el análisis de la riqueza se utilizó una distribución de errores de Poisson y una función de enlace log mediante la función `glmer()` del paquete `lme4` (Bates et al., 2015). Sin embargo, los datos de densidad y biomasa presentaban una elevada cantidad de ceros, por lo que se analizaron mediante la función `glmmTMB()` del paquete `glmmTMB` (Brooks et al., 2017). Este paquete ha sido recientemente desarrollado y es más flexible que otros paquetes para el análisis de datos inflados en ceros y que puedan presentar sobredispersión, permitiendo incluso el modelado de estos parámetros en base a distintos factores. Además, permite la inclusión de factores aleatorios, los cuales son integrados usando la aproximación de Laplace. Se emplearon distribuciones binomiales negativas y funciones de enlace log. La selección de los modelos se realizó a través del Criterio de Información de Akaike (AIC) y se seleccionó el modelo más parsimonioso (menor valor de AIC; Burnham & Anderson, 2002). La validación de los modelos se realizó a través de una inspección visual de los residuos y mediante tests de Shapiro. Los modelos finales no mostraron desviaciones de la homogeneidad de varianzas y la normalidad. Además se analizó la presencia de outliers y cuando fue conveniente se excluyeron y se repitió el modelo. Por último, se evaluó el ajuste de los modelos finales mediante el cálculo de la R2 marginal mediante la función `r2()` del paquete `sjstats` (Lüdecke, 2018). Todos los análisis fueron realizados en R (R Core Team, 2017). Los resultados mostraron que las reservas más efectivas para la conservación de peces fueron Cabo de Palos y Es Freus, al contrario de lo que ocurre con Menorca, lo cual no concuerda con los mayores niveles de producción primaria, por

lo que otros factores como el grado de vigilancia y el tamaño de la reserva pueden estar explicando los patrones observados. En cuanto al nivel de protección, se observó que de forma general las reservas integrales son las más eficaces para las tres variables respuesta, aunque cuando las reservas fueron analizadas una a una fue la zona de reserva parcial la que mostró valores positivos y estadísticamente significativos.

Paquetes de R que emplea: glmer, glmmTMB

Referencias/Bibliografía

- Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1¿48. doi:10.18637/jss.v067.i01
- Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Mächler, M. & Bolker, B.M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal* 9, 378¿400.
- Burnham, K.P., Anderson, D.R. & Laake, J.L. (1980). Estimation of density from line transect sampling of biological populations. *Wildlife Monographs* 72, 1¿202.
- Lüdecke, D. (2018). sjstats: Statistical Functions for Regression Models (Version 0.16.0). doi: 10.5281/zenodo.1284472, <URL: <https://CRAN.R-project.org/package=sjstats>>.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. (2009). *Mixed effects models and extensions in ecology with R*. Springer, New York, New York, USA.

A302. Analítica predictiva en averías de componentes electrónicos

Aleix Cartró Pérez^a Jaume Puigbò Sanvisens^b

^aInnova-tsn. aleix.cartro@innova-tsn.com

^bInnova-tsn. jaume.puigbo@innova-tsn.com

Cada vez más las compañías utilizan sensores o contadores para poder medir y controlar todo tipo de procesos. Es realmente importante detectar a tiempo las anomalías o fallos lo más rápido posible, e incluso anticiparse a ellos. Habitualmente, con este tipo de problemas tenemos una gran cantidad de datos registrados, ya que estos aparatos almacenan información hasta que tienen algún defecto o error. La mayoría de estos registros se encuentran en un espacio temporal en que el contador funcionaba correctamente. Por lo tanto, como es de suponer, nos encontramos con unos datos totalmente desbalanceados. En la ponencia, trataremos de predecir con una semana de antelación las anomalías de unos contadores electrónicos, de consumo de los clientes, para una compañía del sector energético, con técnicas de machine learning. Vamos a plantear cómo crear el dataset para una clasificación binaria, y comentaremos las peculiaridades de dicho dataset. Veremos una comparativa entre varios modelos usando una métrica de evaluación apropiada para este problema. Así mismo, justificaremos la elección de dicha métrica de evaluación.

A303. Visualización y Monitorización de Servicios con R y Shiny

Jorge Pradas Moscardó

Ante el reto de visualizar el estado de los servicios ofrecidos por Edicom nos planteamos el uso de R como herramienta para poder representar de manera sencilla gran cantidad de estados de nuestros servicios. Por un lado veremos la solución adoptada para comprobar la velocidad de los servicios a través de la librería Gplots (Warnes, 2016) y su función Heatmap. Con Shiny (RStudio)

elaboramos un formulario para poder consultar histórico por servicios. Además veremos como comprobar el estado de las diferentes webapps ofrecidas por un servicio web y visualizados los resultados con Gplots.

Referencias/Bibliografía

- Warnes, G (2016) Various R Programming Tools for Plotting Data. <https://www.rdocumentation.org/packages/gplots/versions/3.0.1>
- Shiny from Rstudio. <https://shiny.rstudio.com/>
- McCandless, David (2010) La información es bella. Integral-RBA Editores

A304. Predicción e interpolación dinámica de los niveles de contaminación en base al tráfico y la dirección del viento

Lidia Contreras-Ochando^a César Ferri^b

^aDepartamento de Sistemas Informáticos y Computación. Universitat Politècnica de València liconoc@dsic.upv.es

^bDepartamento de Sistemas Informáticos y Computación. Universitat Politècnica de València cferri@dsic.upv.es

En este trabajo se presenta un método para predecir e interpolar los niveles de contaminación atmosférica de cuatro contaminantes (NO, NO₂, O₃, SO₂) en la ciudad de Valencia. Se estudian diferentes técnicas para incorporar la dirección del viento en los modelos de predicción y se analiza cómo extrapolar las predicciones a toda la ciudad. Para este propósito, se propone un nuevo método de interpolación que tiene en cuenta la dirección del viento a la hora de calcular el resultado. Los experimentos con validación cruzada muestran que este método mejora los resultados en comparación con otros métodos conocidos. Finalmente se presenta airVLC, una aplicación web para mostrar los niveles de contaminación en la ciudad. Todos los métodos y experimentos de este trabajo han sido generados con R.

Referencias/Bibliografía

- F. Ballester Díez, J. M. Tenías, S. Pérez-Hoyos, Efectos de la contaminación atmosférica sobre la salud: una introducción, Revista española de salud pública 73 (2) (1999) 109–121.
- C. global España, 2020/2050. energía, Economía y Sociedad.
- OMS Calidad del aire (exterior) y salud, <http://www.who.int/mediacentre/factsheets/fs313/es/>.
- WHO WHO Global Urban Ambient Air Pollution Database (update 2016), <http://www.who.int/phe/health topics/outdoorair/databases/cities/>
- A. y. M. A. Ministerio de Agricultura, Informe de la calidad del aire en España 2014, http://www.magrama.gob.es/es/calidad-y-evaluacion-ambiental/temas/atmosfera-y-calidad-del-aire/informeevaluacioncalidadaireespana2014_final_tcm7-398522.pdf (2015).
- R. Bivand, T. Keitt, B. Rowlingson, rgdal: Bindings for the geospatial data abstraction library, R package version 0.8-16.
- D. C. Carslaw, K. Ropkins, openair-an r package for air quality data analysis, Environmental Modelling and Software 27–28 (0) (2012) 52–61. doi:10.1016/j.envsoft.2011.09.008.
- R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (2015). URL <http://www.R-project.org/>
- Z. Reitermanova, Data splitting, WDSs 10 proceedings of contributed papers, Part 1 (2010) 31–36.
- L. Provincias, Enciclopedia Universal Las Provincias, Domenech S.A., 1990.
- D. Shepard, A two-dimensional interpolation function for irregularly-spaced data, in: Proceedings of the 1968 23rd ACM National Conference, ACM '68, ACM, New York, NY, USA, 1968, pp.517–524. doi:10.1145/800186.810616. URL <http://doi.acm.org/10.1145/800186.810616>
- P. J. Ribeiro Jr, P. J. Diggle, geor: a package for geostatistical analysis, R news 1 (2) (2001) 14–18.

- D. Kahle, H. Wickham, ggmap: Spatial visualization with ggplot2., R Journal 5 (1)

Ponencia Invitada I

From the package FactoMineR to a project on exploratory multivariate analysis or how to improve the visibility of its package

Francois Husson^a

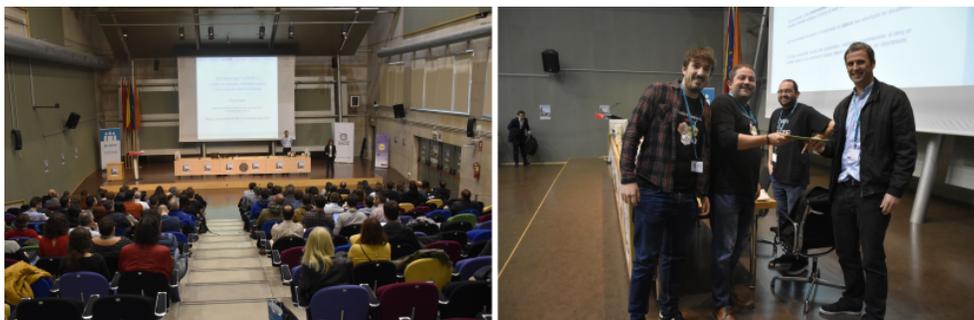
^aDepartment Statistics & Computer science, Agrocampus Ouest, Rennes, Francia, francois.husson@agrocampus-ouest.fr



FRANÇOIS HUSSON

The design of new statistical methods or methodological approaches is more and more often highlighted by the construction of an R package. The creation of such a package makes it possible to share its work with the whole scientific community by facilitating in particular the comparison of methods. This package creation is time-consuming and it would be a pity that the time spent is not beneficial for either the package designer or the community. Now the multiplication of packages, on the CRAN or on other repositories like Github, complicates the distribution of its package and limits its visibility. Many packages are thus unused... and therefore unfortunately useless. We will discuss in this presentation various actions and vectors of diffusion allowing to improve the visibility and the accessibility of its package to the potentially interested users. Some actions can be easily performed and could (should?) be implemented by any package designer. Other actions that are more complex and take longer to implement can be built over the long term and help to gradually build distance learning from simple tutorials to the creation of a MOOC. We will take the example of the FactoMineR package that is dedicated to exploratory multivariate analysis methods such as PCA, multiple correspondence analysis or clustering. We will see its evolution from a simple package submitted for the first time in April 2006 to a more global project on exploratory multivariate analysis. Indeed, there are now complementary packages that allow the construction of interactive graphics, the management of missing data or the generation of automatic interpretation of results. Moreover, there are additional materials like a website, a Google group to ask questions, research papers, books (in French, English or Spanish) and Youtube tutorials. We will also see that the FactoMineR package is at the heart of a MOOC on exploratory multivariate analysis available in French or English. Already attended by more than 25 000 participants (with 4 sessions in French and two in English), this MOOC participates in the dissemination of methods and research activities in the business world and in the academic world.

Vidéo de la Conferencia plenaria del Dr.-François Husson: <https://tv.um.es/video?id=124941f>.



Comunicaciones B

Cuadro 4.4: **Ponencias Salón B** (id es un enlace a la presentación).

id	Título	Autorías	Entidad
B01	rjazz, an R package to control Jazz, the lightweight data processing framework	Santiago Basaldúa	BBVA Data & Analytics
B02	Time series model selection from a bayesian network practical approach	Ana Arias Botey, Gabriel Valverde Castilla	CUNEF
B03	Desarrollo de una aplicación web para el conjuntos de datos de impactron y Post-AWS	Andrés Bishop y Enric Aguilar	Centre for Climate Change, Univ. Rovira i Virgili
B04	Comparación de modelos espacio-temporales de datos de panel con R	Patricia Carracedo y Ana Debón	Univ. Internacional de Valencia y Univ. Politécnica de Valencia
B05	Anticipar el protocolo de contaminación usando el portal de datos abiertos Madrid	Marta López Ron, María Neira Bustamante, Fernando Sebastián Huerta	INNOVA-TSN
B06	Text Mining para mejorar nuestro Customer Relationship Management (CRM)	Carlos Vecina Tebar y Pablo Cánovas	Gnarum Tecnología y Energía

B01. rjazz, an R package to control Jazz, the lightweight data processing framework.

Santiago Basaldúa^a

^aBBVA Data & Analytics, santiago@dybot.com

The talk introduces the package rjazz, (in cran). Jazz (the server <https://github.com/bbvadata/jazz-server>) and rjazz (the client <https://cran.r-project.org/web/packages/rjazz/index.html>) were released as OSS by BBVA end of 2017. This corresponds to the current version actually available at cran. During 2018, a community of volunteers started refactoring Jazz into the product showed in this talk. <https://github.com/kaalam/Jazz> It will be available before Nov '18 including:

From 20,000 feet: Before a detailed introduction, in case you feel the need to put a label on Jazz or want to pitch it in 10 seconds, you may ask.

Is it a distributed no-sql database, a language, an http server, a deep learning framework Jazz is a new kind of data product. It combines all these things and some others (an automated creator of APIs, an abstraction over data sources) while being just one extremely lightweight process. It is free software created on top of third party free software. It is written in C/C++, available through R and Python packages and also through a REST API. The framework is Linux only, but, since it is an http back-end, Jazz consumers include all devices and operating systems. Jazz is designed for millions of users interacting concurrently with large clusters.

What makes it special?

1. Blocks are volatile, persisted, distributed or ubiquitous as defined. Abstraction from extremely small to extremely large is automatic. The same function works on a small vector or a distributed data-frame.
2. Extremely efficient persistence (a memory mapped file) in the same process allows sampling without copying. (No need of a DB)
3. Extremely efficient REST API in the same process with a thread pool. (No need of an http server)
4. A full language that is close to perfect performance (assuming enough tensor size), thread safe and auto distributing instead of just some optimized data mangling primitives. (No need of an interpreter)
5. Distributed data-frames where cells are tensors (E.g., video or sound tracks, not links to files)
6. An abstraction over the file system that allows reading/writing binary and compressed files on demand.

4. RESÚMENES DE PONENCIAS

7. Tree Search as a first class citizen
8. Reinforcement learning “out of the box”
9. Inference on small samples done efficiently
10. Support for automatic programming at bytecode level
11. Deep Learning as a first class citizen

What will the R API look like: APIs And finally, the API. Thanks for reading this far! The API is a keepr named / (root in Unix) where all the keepsrs and blocks are linked. The API is the same for all the languages including REST (where root is // since just one / is used in normal urls.) Any API call can be seen as an assignment. In an assignment, X=Y there are two sides, the left side (X) called lvalue and the right side (Y) called rvalue. This is C++ naming.

In Jazz an lvalue is one of:

- A chain of keepsrs abstracted as block ids starting from root ending with the name of a block. They must all exist except, possibly, the last name. If the last name is new, it is created, if it exists, overridden.
- A data block that will be returned as a Python object, R object or http (GET) resource. In Jazz an rvalue is one of:
 - A block constructor. A constant expression that can be used to build a block from it.
 - Chains of keepsrs abstracted as existing blocks starting from root. This includes functions and blocks passed as arguments to functions.
 - A combination of the previous two.
- A data block that will be passed as a Python object, R object or http (PUT) resource.
- A delete predicate. This deletes the corresponding lvalue. Since the API has to be REST compatible and is intended for using over a network.
- All rvalue evaluations are safe. They cannot have side effects. Function calls cannot

Paquetes de R que emplea: rjazz

B02. Time series model selection from a bayessian network practical approach

Ana Arias Botey^a Gabriel Valverde Castilla^b

^aCUNEF ana.arias@cunef.edu

^bCUNEF gabriel.valverde@cunef.edu

The goal of this essay is to determine the model that best predicts a time series based on its prior characteristics. To address this problem we’ve used a process that includes the transformation of a table of events into time series, the prediction obtained through different models and the analysis of the results obtained by each of them. The data has been collected from Makridakis Competitions 4 (also known as the M Competitions). This dataset is a sample of 100000 time series of different types such as demographic, financial, industrial, micro, macro and others has been used with different granularities also, such as, annual, quarterly, monthly, weekly, daily and hourly. The models that have been used are based on the decomposition of a time series in its components of state (ETS, BSTS and Prophet) and in the Box Jenkins methodology (ARIMA). Finally, given the characteristics of a time series, in order to choose the most efficient model, a Bayesian network has been used to determine the probability of the model that predicts it with a minimum error range. The programming language used during the making of this essay has been R.

Paquetes de R que emplea: readr, TSPred, plotly, data.table, dplyr, plotrix, RColorBrewer, forecast, psych, prophet, bst, tseries, zoo, lubridate, stats, TSA, tictoc, rlist, magrittr, cluster, factoextra, fpc, parallel, rlist, magrittr, xts, ggplot2, rpart, pipeR, shiny, shinydashboard, shinyloaders, shinycss, networkD3, visNetwork, bnlearn, sparklyr, tidyverse, yaml, rlang

B03. Desarrollo de una aplicación web para el conjuntos de datos de impactron y post-aws

Andrés Bishop^a Enric Aguilar^b

^aC3 Centre for Climate Change, Universidad Rovira i Virgili, andres.bishop@urv.cat

^bC3 Centre for Climate Change, Universidad Rovira i Virgili

Los conjuntos de datos climáticos largos, ininterrumpidos y homogéneos son esenciales para la investigación climatológica. Las aplicaciones y los servicios de usuario conforman grandes conjuntos de datos que generalmente no son fácilmente accesibles. En este trabajo se presenta POSTIMPACT, una aplicación web Shiny que permite a los usuarios interactuar y visualizar de diversas formas con datos climatológicos de las redes IMPACTRON y POST-AWS de manera ubicua, directa y flexible, fomentando la colaboración entre investigadores y partes interesadas en general. La aplicación ha sido probada con éxito realizando el análisis de algunos estadísticos, específicamente el comportamiento de la media tx en el Observatorio Ebro, donde se encontró que 8 segmentos formaban tres grupos según las características de sus sesgos, y donde la magnitud de las diferencias de medias de estos grupos varía a lo largo del año, con los meses de invierno y otoño mostrando las diferencias medias mínimas. Además, el análisis de la media de tx y la media de tn para todo el conjunto de datos de IMPACTRON mostró un sesgo caracterizado por un ligero signo negativo para tn, debido, en esencia, al signo negativo de las diferencias de las medias de primavera y verano. Trabajo de fin de grado. Grado de Geografía y Ordenación del Territorio. Universidad Rovira i Virgili. Tutor: Dr.Enric Aguilar. Se puede acceder a POSTIMPACT en: <https://bishopf.shinyapps.io/Impactron/>

Referencias/Bibliografía:

- E. Aguilar, I. Auer, M. Brunet, T. C. Peterson, J. Wieringa, GUIDELINES ON CLIMATE METADATA AND HOMOGENIZATION, Tech. rep., World Meteorological Organization, Geneva (2003).
- O. Baddour, H. Kontongomde, GUIDELINES FOR MANAGING CHANGES IN CLIMATE, Environment (62). B. Bahar, A. F. Tuncel, E. W. Holmes, D. T. Holmes, An interactive website for analytical method comparison and bias estimation, Clinical Biochemistry 50 (18) (2017) 1025–1029. doi:10.1016/j.clinbiochem.2017.08.008.
- M. Brunet, J. Asin, J. Sigró, M. Bañón, F. García, E. Aguilar, J. E. Palenzuela, T. C. Peterson, P. Jones, The minimization of the screen bias from ancient Western Mediterranean air temperature records: An exploratory statistical analysis, International Journal of Climatology (31). doi:10.1002/joc.2192.
- W. Chang, J. Cheng, J. Allaire, Y. Xie, J. McPherson, Shiny - Web Application Framework for R. R Package Version (2017). URL <https://cran.r-project.org/web/packages/shiny/index.html>
- J. Cheng, Y. Xie, B. Karambelkar, leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library (2018). URL <https://cran.r-project.org/package=leaflet>
- X. Feng, M. C. Castro, E. Linde, M. Pape, Armadillo Mapper, Tropical Conservation Science 10 (2017) 194008291772413. doi:10.1177/1940082917724133.
- A. Gilabert, Assessment of the bias introduced by the automatization of climate records combining climatological and metrological approaches, Ph.D.thesis, Rovira i Virgili (2016).
- T. Klein, A. Samourkasidis, I. N. Athanasiadis, G. Bellocchi, P. Calanca, webXTREME: R-based web tool for calculating agroclimatic indices of extreme events, Computers and Electronics in Agriculture 136 (2017) 111–116. doi:10.1016/j.compag.2017.03.002.
- P. Moraga, SpatialEpiApp: A Shiny web application for the analysis of spatial and spatio-temporal disease data, Spatial and Spatio-temporal Epidemiology 23 (2017) 47–57. doi:10.1016/j.sste.2017.08.001.
- B. M. Zaragoza, J. Torres-Prieto, A. Belda-Antolí, A. Rosado-Abad, C. Cortés-Samper, A. Ramón-Morte, Small data para educación: más de un millón de pirámides de población de España, XVII Congreso Nacional de Tecnologías de Información Geográfica (2016) 454–462.

B04. Comparación de modelos espacio-temporales de datos de panel con R

Patricia Carracedo^a Ana Debón^b

^aUniversidad Internacional de Valencia patricia.carracedo@campusviu.es

^bUniversitat Politècnica de València

Actualmente, los países disponen de información sobre la mortalidad a lo largo del tiempo y desagregada por edades. Por ello, sería interesante utilizar esta información para estudiar la dependencia espacial de la mortalidad europea en edades de jubilación a lo largo del tiempo e identificar que variables socioeconómicas afectan a dicha mortalidad mediante un modelo econométrico espacio-temporal. Como es importante tener en cuenta la variable sexo, este estudio diferencia entre sexo masculino y femenino. De esta forma se pueden identificar patrones de comportamiento diferentes en mortalidad. Por todo ello, el objetivo de este trabajo es evaluar y seleccionar el mejor modelo espacio-temporal para cada sexo que se ajuste a los datos utilizando el software estadístico R. Los modelos utilizados tienen en cuenta la dimensión espacial y temporal del panel de datos utilizado. Este panel corresponde a datos de mortalidad de hombres y mujeres jubilados con edades comprendidas entre 65 y 110 años, 26 países europeos durante el período 1995-2009. Además, para estos 26 países y 15 años considerados en el estudio se ha obtenido información de 4 covariables: tasa de crecimiento del Producto Interior Bruto (PIB), gasto sanitario, CO₂ y gasto en educación. Los modelos obtenidos fueron validados analizando sus residuos y se compararon mediante medidas de bondad de ajuste como el coeficiente de determinación (R²), criterio de información de Akaike (AIC) y criterio de información bayesiano (BIC). Concluyendo el modelo que mejor ajusta para los hombres es el modelo de retardo espacial con efecto fijo espacial y temporal donde las 4 covariables son significativas, mientras que para mujeres el mejor fue el modelo con efecto fijo espacial y temporal donde las covariables tasa PIB y gasto educación son significativas.

Paquetes de R que emplea: splm, plm, spdep, HMDHFDplus

Referencias/Bibliografía

- Anselin, L. (1988). Spatial econometrics: Methods and models. Kluwer Academic. Boston.
- Anselin, L. (1995). Local Indicators of Spatial Association-LISA. *Geographical Analysis*, 27(2):93-115.
- Bivand, R. (2014). spdep: Spatial dependence: weighting schemes, statistics and models. R package, version 0.5-77. <http://CRAN.R-project.org/package=spdep>.
- Breusch, T. and Pagan, A. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies*, 97:239-253. Burridge, P. (1980). On the Cliff-Ord test for spatial autocorrelation among regression residuals. *Geographical Analysis*, 4:267-284.
- Croissant, Y., Millo, G., et al. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software*, 27(2):1-43.
- Elhorst, J. (2014). Spatial econometrics: from cross-sectional data to spatial panels. Springer.
- Hsiao, C. (2014). Analysis of Panel Data (3rd edition), chapter Simple regression with variable intercepts. Cambridge University Press, Cambridge.
- Human Mortality Database (2016). University of California, Berkeley (USA), and Max Planck Institute for - Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 12th July 2016).
- Julious, S., Nicholl, J., and George, S. (2001). Why do we continue to use standardized mortality ratios for small area comparisons. *Journal of Public Health*, 23(1):40-46. Kennedy, P. (2003). A guide to econometrics. MIT press. Cambridge, MA.
- Millo, G. and Piras, G. (2012). splm: Spatial panel data models in R. *Journal of Statistical Software*, 47(1):1-38.
- Moran, P. A. (1950a). Notes on continuous stochastic phenomena. *Biometrika*, pages 17-23.
- Moran, P. A. (1950b). A test for the serial independence of residuals. *Biometrika*, pages 178-181.
- Riffe, T. (2015). Reading human fertility database and human mortality database data into R. Technical Report TR-2015-004, MPIDR.
- The World Bank Database (2018). World Development Indicators. Data download on 6th March 2018 en <http://data.worldbank.org/>.

Esta investigación fue financiada por la subvención Ignacio H. de Larramendi de la Fundación Mapfre.

B05. Anticipar el protocolo de contaminación usando el portal de datos abiertos Madrid

Marta Lopez Ron ^a María Neira Bustamante ^b Fernando Sebastián Huerta ^c

^aINNOVA-TSN fernando.sebastian@innova-tsn.com

^bINNOVA-TSN maria.neira@innova-tsn.com

^cINNOVA-TSN marta.lopez@innova-tsn.com

El problema de la contaminación afecta cada vez a más ciudades en el mundo. Por ello es necesario disponer de una herramienta que posibilite actuar a tiempo, antes de superar los niveles perjudiciales para la salud. En la reciente KDD CUP 2018, competición internacional más importante en el campo del Data Mining, el objetivo fue la predicción de la calidad del aire para las 48 horas siguientes en las estaciones de Londres y Beijing. El objetivo de la charla es mostrar, mediante técnicas de Machine Learning y lenguaje R, el reto de mejorar la calidad del aire de Madrid y el servicio al ciudadano, y las conclusiones que hemos observado.

- Diseñar un modelo predictivo con datos obtenidos gracias al portal Open Data del consistorio madrileño.
- Mejorar la predicción calidad aire (NO₂) permitiendo alertar a la ciudadanía sobre la aplicación de los protocolos de contaminación con más de un día de antelación.

B06. Agrupamiento con restricciones para configurar sesiones en conferencias científicas con R + Shiny

Cèsar Ferri ^a Diego Fernando Vallejo Huang ^b Paulina Adriana Morillo Alcívar ^c

^aUniversitat Politècnica de València cferri@dsic.upv.es

^bUniversidad San Francisco de Quito

^cUniversidad Politecnica Salesiana

Distributing papers into sessions in scientific conferences is a task consisting in grouping papers with common topics and considering the size restrictions imposed by the conference schedule. This problem can be seen as a semi-supervised clustering of scientific papers based on their features. This work presents a web tool called ADoCS that solves the problem of configuring conference schedules by an automatic clustering of articles by similarity using a new algorithm considering size constraints. The system tool has been implemented totally in R. Specifically, we have employed the following R libraries:

- tm: The tm package which contains a set of functions for text mining in R. In this project we have use the functionalities related to text transformations (stemming, stopwords, TF-IDF, ...) with the English corpora.
- wordcloud: The wordcloud package includes functions to show the most common terms in a vocabulary in form of the popular wordcloud plot.
- proxy: The proxy package allows to compute distances between instances. The package contains implementations of popular distances, such as: Euclidean, Manhattan or Minkowski, etc.
- Rglpk: The Rglpk package represents an R interface to the GNU Linear Programming Kit. GLPK is a popular open source kit for solving linear programming (LP), mixed integer linear programming (MILP) and other related problems with optimisation.

The ADoCS source code and some datasets about conferences to test the tool can be found in <https://github.com/dievalhu/ADoCS>.

The graphical user interface has been developed by means of the Shiny package. This package constitutes a framework for building graphical applications in R. Shiny is a powerful package for converting basic R scripts into interactive web applications without requiring programming web knowledge. As a result we obtain an interactive web application that can be executed locally using a web browser, or can be uploaded to a shiny application server where the tool is available for a general use. In this case, we have uploaded the application to the free server <http://www.shinyapps.io/>.

Paquetes de R que emplea: Shiny, tm, Rglpk

Referencias/Bibliografía

- D. F. Vallejo Huanga, P. A. Morillo Alcívar, C. Ferri Ramírez, Adocs: Automatic designer of conference schedules, in: Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics, Association for Computational Linguistics, 2017, pp.41–44. URL <http://aclweb.org/anthology/E17-3011>
- D. Vallejo-Huanga, P. Morillo, C. Ferri, Semi-supervised clustering algorithms for grouping scientific articles, *Procedia Computer Science* 108 (2017) 325 – 334, international Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland. doi:<https://doi.org/10.1016/j.procs.2017.05.206>. URL <http://www.sciencedirect.com/science/article/pii/S1877050917307998>

B07. Text Mining para mejorar nuestro Customer Relationship Management (CRM)

Carlos Vecina Tebar^a

^aData Scientist en Gnarum Tecnología y Energía carlosvecinatebar@gmail.com

Después de una breve introducción a los conceptos clave de la minería de textos, a la metodología uima y al estado del arte en algoritmos como word2vec, nos centraremos principalmente en las posibilidades que nos brindan en el campo de la empresa y las finanzas en general y del CRM en particular. Como caso de uso de ejemplo finTech, mezclando un enfoque supervisado y no supervisado, extraeremos insights de datos abiertos sobre usuarios de aplicaciones bancarias, tanto de banca tradicional como nuevos competidores. El tiempo dedicado a una y otra parte podrá balancear dependiendo del perfil del auditorio y del contenido del resto de charlas.

Paquetes de R que emplea: Tm, snowball, dplyr, caret, keras, xgboost

Comunicaciones C1

Cuadro 4.5: **Ponencias Salón C1** (id es un enlace a la presentación)

id	Título	Autorías	Entidad
C101	RSGIS, una aplicación shiny basada en Sistemas de Información Geográfica	José Antonio Sánchez Martí	Univ. de Murcia
C102	Structural Equation Modeling with lavaan to explain the effects of co-creation on consumer behavior	Rocío Alarcón-López, Inés López-López and Salvador Ruiz-de-Maya	Univ. de Murcia
C103	deaR. Data Envelopment Analysis con R	Vicente Coll Serrano, Rafael Benítez Suárez y Vicente J. Bolós Lacave	Univ. de Valencia
C104	Análisis de la influencia de amigos en Facebook con R	Manuela López, Mariola Palazón, María Sicilia	Univ. de Murcia
C105	Applying Data Science to Fraud Prevention and Detection: SALER Analytics	Fernando Martínez-Plumed, Juan Carlos Casamayor, César Ferri, Jon Ander Gómez y Eduardo Vendrell	Univ. Politécnica de Valencia
C106	Agrupamiento con restricciones para configurar sesiones en conferencias científicas con R + Shiny	César Ferri, Diego Fernando Vallejo Huang y Paulina Adriana Morillo Alcívar	UPV, Univ. San Francisco de Quito y Univ. Politécnica Salesiana

C101. RSGIS, una aplicación shiny basada en Sistemas de Información Geográfica

Jose Antonio Sánchez Martí^a

^aUniversidad de Murcia, joseantonio.sanchez7@um.es

La finalidad del software RSGIS es crear una herramienta básica y online para trabajar con Sistemas de información Geográfica, siendo accesible desde cualquier dispositivo y sin la necesidad de instalar ningún programa. La principal función del software es: por un lado, visualizar de forma interactiva polígonos de ficheros espaciales; y por otro lado, la representación de puntos mediante sus coordenadas. Es necesario en primer lugar, cargar como base un fichero shapefile (siendo necesario los ficheros con extensiones shp, dbf, shx y prj) antes de añadir coordenadas (en formato csv). Una vez cargado un shapefile, podemos interactuar con la base de datos asociada al fichero espacial. El software también dispone de una opción para editar los datos y puntos. Las variables asociadas a los datos espaciales pueden ser filtradas según un cierto valor. A su vez, los puntos pueden editarse según el tamaño, color y transparencia. La siguiente opción del software es clave, hablamos de la selección de las variables que posteriormente utilizaremos en el análisis de los datos. Se puede seleccionar una o varias variables según el tipo de análisis a realizar. RSGIS nos permite analizar los datos de dos formas distintas, una mediante la elaboración de mapas temáticos según el nivel de una variable, y otra mediante el análisis de conglomerados. En este último, tenemos las opciones disponibles para la creación de los conglomerados, desde los métodos de las distancias y las aglomeraciones a emplear, hasta el número de conglomerados. Los conglomerados realizados serán añadidos a la base de datos y puedes descargarlos para su posterior uso. También en dicha sección, encontramos una opción que nos muestra un resumen de estadísticas descriptivas de cada conglomerado. Los mapas elaborados tanto por niveles, como por conglomerados pueden descargarse en formato pdf desde la pestaña inicial. Adicionalmente, tenemos incorporada una pestaña dedicada a la creación de rutas según una geoposición dada, de gran utilidad al trabajar con puntos. Para terminar con las funciones de la aplicación web, tiene incluidos tres botones: un manual de ayuda al usuario, un botón para reiniciar la aplicación y otro para contacto. Las principales librerías con que se ha escrito el código son: shiny, leaflet, rgdal, ggmap, rmarkdown. En definitiva, disponemos a nuestro alcance de una herramienta rápida y práctica, que nos permite realizar análisis exploratorios del territorio. Destacando que podemos interactuar y descargar los mapas creados para apoyar nuestras investigaciones.

C102. Structural Equation Modeling with lavaan to explain the effects of co-creation on consumer behavior

Rocío Alarcón-López^a, Inés López-López^b and Salvador Ruiz-de-Maya^c

^aUniversidad de Murcia, rocio.alarcon@um.es

^bUniversidad de Murcia

^cUniversidad de Murcia

Continuous advances in technologies along with the Internet have facilitated co-creation, an innovative strategy for marketers aiming to better suit consumers specific needs (Ostrom et al., 2015). A well-known example of co-creation is the ‘Do Us A Flavor campaign, which allows fans to win \$1,000,000 if the potato chips flavor they suggest receives the highest number of votes from other consumers. The goal of this study is to analyze the effects of failed (vs.successful) value co-creation experiences on consumer responses. Co-creation literature assumes that both parties benefit from the interaction and that co-creation is collaborative. However, since not even the most consumer-oriented companies can ensure a hundred percent error-free service, failures could also show up in co-creation activities. The failure context serves as a reference point for consumers’ judgments, by the characteristics of this performance disruption. In these negative experience situations, firms usually develop actions (service recovery strategies) to fix the problem, which may be an opportunity to enhance the relationship but also to worsen it (Gronroos, 1988). The review of the literature on consumer emotions, engagement and attributions took us to propose four hypotheses that positively relate successful co-creation with more positive emotions, satisfaction, engagement and internal attributions when compared to situations of failed co-creation. The methodology is based on a between-subjects experimental design with three co-creation outcome conditions (successful vs.failed vs.control condition). 144 undergraduate students were randomly assigned to one of the three scenarios in an online survey. In the designed context, subjects were asked to come up with a new croquette for a fictitious restaurant. After completing the co-creation task, participants reported their satisfaction with the co-creation outcome, emotions felt and engagement with the restaurant. They also rated attribution, service performance, realism of the scenario and the perceived level of value co-creation. Finally, they rated product knowledge and the questionnaire ended with demographic questions. All dependent variables were assessed using 11-point scales. We analyzed the data on R and we tested our hypotheses using packages *phia* (Martínez, 2015) and *lavaan* (Rosseel, 2012). Moreover, we estimated the structural equation model using bootstrapping.

An index for emotions was calculated where negative emotions were subtracted from positive emotions. Results from analysis of covariance revealed a main effect of the co-creation outcome on emotions, satisfaction with the outcome, engagement and attribution, which confirm our four hypotheses. Simple effects were analyzed with the R package *phia*. In addition to the ANCOVA analyses, we used maximum likelihood estimation SEM with 10,000 bootstrapped samples with the R package *lavaan*. The model showed a good fit to the data and results confirmed that the total effect of failed co-creation outcome via emotions on satisfaction was significant, whereas the total effect of failed co-creation outcome on consumer engagement was not. On the other hand, while successful co-creation has no positive effect on satisfaction, it has a positive impact on consumer engagement with the brand. These findings allow us to conclude that the implementation of co-creation processes can increase consumer engagement with the brand, which goes beyond purchase behavior.

Paquetes de R que emplea: *phia*, *lavaan*

Referencias/Bibliografía

- Gronroos, C. (1988). Service quality: The six criteria of good perceived service. *Review of business*, 9(3), 10.
- Martínez, H. D. R. (2015). Analysing interactions of fitted models. Retrieved from <http://cran.wustl.edu/web/packages/phia/vignettes/phia.pdf>
- Ostrom, A. L., Parasuraman, A., Bowen, D. E., Patricio, L., Voss, C. A., & Lemon, K. (2015). Service research priorities in a rapidly changing context. *Journal of Service Research*, 18(2), 127-159.
- Rosseel Y. (2012). *lavaan*: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36

C103. deaR. Data Envelopment Analysis con R.

Vicente Coll Serrano^a . Rafael Benítez Suárez^b . Vicente J. Bolós Lacave^c .

^aUniversidad de Valenciavicente.coll@uv.es

^bUniversidad de Valencia

^cUniversidad de Valencia

En esta comunicación presentamos un nuevo y versátil paquete de R para efectuar análisis de eficiencia y productividad basado en la metodología DEA (Data Envelopment Analysis). Con este paquete, que hemos denominado deaR, es posible ejecutar tanto modelos DEA convencionales o clásicos (CCR, BCC, índices de Malmquist, eficiencia cruzada, modelos de super eficiencia, etc.) como modelos DEA fuzzy si se trabaja con datos inciertos. deaR está pensado para que pueda ser utilizado tanto por investigadores como usuarios de DEA. Además, deaR dispone también de una versión Shiny que permite que usuarios no familiarizados con R puedan usar fácilmente el programa. Por último, deaR cuenta con un importante número de datasets procedentes de artículos científicos que son utilizados para replicar los resultados alcanzados por los autores en sus respectivos trabajos. Este banco de datos pretende ser una herramienta para facilitar la enseñanza y aprendizaje de la metodología DEA.

C104. Análisis de la influencia de amigos en Facebook con R

Manuela López^a , Mariola Palazón^b y María Sicilia^{c d}

^aDepartamento de Comercialización e Investigación de Mercados, Universidad de Murcia, manuela.lopez@um.es

^bDepartamento de Comercialización e Investigación de Mercados, Universidad de Murcia

^cDepartamento de Comercialización e Investigación de Mercados, Universidad de Murcia

^dDepartamento de Comercialización e Investigación de Mercados, Universidad de Murcia

R es una herramienta cada vez más utilizada en las investigaciones de marketing. Su versatilidad para aplicar diversas estrategias de análisis y su gran potencialidad gráfica ha favorecido su utilización en esta área de conocimiento. El presente trabajo utiliza esta herramienta para el análisis de datos de una investigación en el ámbito de las redes sociales, concretamente, en Facebook. El objetivo es analizar el efecto que tiene sobre un individuo la información que se comparte en las redes sociales sobre marcas y/o productos. Concretamente, se analiza si la similitud entre los individuos que comparten la información, y la fortaleza de la relación que existe entre ambos determinan la influencia que dicha información va a tener en su comportamiento en redes sociales. Además, teniendo en cuenta que la información que nos llega de nuestros contactos en Facebook nos puede llegar de distintas formas con distinto grado de intencionalidad (a través de un mero “me gusta” o a través de la opción más directa “compartir en la bio de un amigo”), proponemos que la intencionalidad a la hora de compartir información moderará las relaciones propuestas. Los datos utilizados en esta investigación fueron recogidos a través de encuestas online. Se realizó un diseño experimental 2x2 entre sujetos en el que se manipularon la intencionalidad que tiene la persona al compartir la información (influencia intencionada vs.no intencionada) y la fortaleza de la relación entre quien recibe la información y quien la comparte (amigo vs.conocido). Para analizar los datos, utilizamos en primer lugar el paquete `xlxs` para la lectura de la base de datos. Con el objetivo de evaluar la fiabilidad y validez de las variables utilizadas se llevó a cabo un análisis factorial confirmatorio (CFA) con el paquete `lavaan`. Para los análisis descriptivos de cada una de las variables y los análisis de regresión utilizamos `psych`. En concreto, los análisis de regresión múltiple los evaluamos con la función `lm`. De los resultados obtenidos destacamos que tanto la fortaleza de la relación como la similitud entre los individuos afectan positivamente al comportamiento del individuo que recibe la información (medido a través de la utilidad de la información compartida en redes sociales y del comercio social). Además, se analizó el efecto moderador de la intencionalidad en las relaciones previamente planteadas, para lo cual se calcula la región de Johnson y Neyman a través de la función `sim_slopes` y su consiguiente representación gráfica. Esta función forma parte del paquete `jttools`. Este análisis nos ha permitido saber que la similitud entre los individuos tiene un efecto positivo sobre el comercio social para cualquier valor de la intencionalidad, aunque dicho efecto disminuye conforme se incrementa la percepción del consumidor respecto a la intencionalidad percibida en el acto de compartir la información del producto/marca. Por otro lado, la fortaleza de la relación entre los individuos solo afecta positivamente a la utilidad percibida de la información cuando el receptor de la información perciba una alta intencionalidad en el acto de compartir la información (8 en una escala de 1 a 10). Cuanto más fuerte sea la relación, más útil percibirá el consumidor la información compartida. Con este estudio

ponemos de manifiesto la utilidad de la herramienta R en las investigaciones en marketing, tanto desde el punto de vista analítico como gráfico.

Paquetes de R que emplea: xlxs, lavaan, psych, jtools

C105. Applying Data Science to Fraud Prevention and Detection: SALER Analytics

Fernando Martínez-Plumed^a, Juan Carlos Casamayor^b, Cèsar Ferri^c, Jon Ander Gómez^d y Eduardo Vendrell^e

^aUniversitat Politècnica de València, fmartinez@dsic.upv.es

^bUniversitat Politècnica de València

^cUniversitat Politècnica de València

^dUniversitat Politècnica de València

^eUniversitat Politècnica de València

With the aim of detecting and preventing bad practices and fraud in public administration, in this work we present SALER (“Rapid Alert System”, called after the Spanish initials), a data science-based solution which has been developed in cooperation between the Universitat Politècnica de València (research and innovation) and Generalitat Valenciana (governance and public administration). By making better use of relevant information of public expenditure and the expertise available within the public institutions, SALER seeks to perform a quantitative and qualitative analysis of this information based on requirements and specifications set by the audit, control and governing bodies in the city of Valencia, while ensuring better compliance with the policies, rules and regulations. Specifically, SALER is being designed for analysing and responding to a series of specific questions and requisites regarding fraud and corruption. To do so, we perform intelligent mining of internal and external data sources from administrative procedures by developing data analysis algorithms that cross-check data from various public and private institutions. Machine learning, pattern detection, data analysis and intelligent mining, as well as other procurement monitoring and analytic techniques are being used to try to identify projects susceptible to risks of fraud, conflict of interests or irregularities. SALER is being developed under the framework of the Valencian law “on the general inspection of services and the alert system for the prevention of bad practices in the administration and public sector” 1, which aim is to bring important advances in preventing and combating corruption and fraud in public administration, as well as provide legal support and security to the development of the SALER project in terms of data, means and instruments.

Paquetes de R que emplea: dplyr, tidyr, shiny, caret, ggplot

Referencias/Bibliografía

- Winston Chang, Joe Cheng, J.J. Allaire, Yihui Xie, and Jonathan McPherson. shiny: Web Application Frame-work for R, 2017. R package version 1.1.0 - For new features, see the 'Changelog' file (in the package source).
- Max Kuhn. caret: Misc functions for training and plotting classification and regression models, 2018. R package version 6.0.8 - For new features, see the 'Changelog' file (in the package source).
- Fernando Martnez-Plumed, Juan Carlos Casamayor, Csar Ferri, Jon Ander Gmez, and Eduardo Vendrell. SALER: a data science solution to detect and prevent corruption in public administration. In Proceedings of the 3rd International Workshop on Data Science for Social Good (SoGood 2018 @ ECML 2018), Dublin, Ireland, 2018.
- R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2018.
- Ian Robinson, Jim Webber, and Emil Eifrem. Graph databases: new opportunities for connected data. ” O’Reilly Media, Inc.”, 2015.
- Jim Webber. A programmatic introduction to neo4j. In Proceedings of the 3rd annual conference on Systems, programming, and applications: software for humanity, pages 217–218. ACM, 2012.
- Hadley Wickham. dplyr: A Grammar of Data Manipulation, 2018. R package version 0.7.6 - For new features, see the 'Changelog' file (in the package source).

- Hadley Wickham. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*, 2018. R package version 3.0.0 - For new features, see the 'Changelog' file (in the package source).
- Hadley Wickham. *tidyr: Easily Tidy Data with 'spread()' and 'gather()' Functions*, 2018. R package version 0.8.1 - For new features, see the 'Changelog' file (in the package source).

Comunicaciones C2

Cuadro 4.6: **Ponencias Salón C2** (id es un enlace a la presentación)

id	Título	Autorías	Entidad
C201	Desarrollo de materiales y aplicaciones científicas interactivas en entorno web para la resolución de problemas de optimización	Manuel Muñoz Márquez	Univ. de Cádiz
C202	Explorando los datos públicos de España: un paquete para navegar la API de datos.gob.es	Jorge Cimentada, Jorge López	Univ. Pompeu Fabra, Kernel Analytics Bringing
C203	Standards to Quality Control with R	Manuel Alfaro García, Emilio López Cano	Univ. de Castilla-La Mancha
C204	The efficiency of public lighting service in Spain	Pedro-José Martínez Córdoba, María-Dolores Guillamón López, Bernardino Benito López, Úrsula Faura Martínez	Univ. de Murcia
C205	LAU2boundaries4spain: un paquete de R con contornos municipales para el periodo 2002-2018	Pedro J. Pérez, Francisco Goerlich	Univ. de Valencia e IVIE

C201. Desarrollo de materiales y aplicaciones científicas interactivas en entorno web para la resolución de problemas de optimización

Manuel Muñoz Márquez^a

^aUniversidad de Cádiz, manuel.munoz@uca.es

Se presenta una experiencia en la que el alumnado del Grado en Matemáticas elaboran aplicaciones interactivas en línea para la resolución de problemas de optimización como trabajos en la asignatura de Programación Matemática.

La realización de aplicaciones interactivas en línea está fuera del alcance de cualquier programador novel. Esto hace pensar que es inviable que el alumnado del Grado en Matemáticas pueda abordar el desarrollo de una aplicación que resuelva un problema real y que además lo haga de forma interactiva y en entorno web. La metodología desarrollada, y que se presenta aquí, muestra como partiendo del entorno de desarrollo creado por los docentes, el alumnado alcanza el objetivo de desarrollar una aplicación en línea interactiva que resuelve un problema real de optimización sin más límite que la capacidad del ordenador donde se ejecuta el software. Además, todo este proceso se realiza de forma que muchos de los aspectos importantes en el desarrollo de cualquier aplicación de este tipo y que no aportan formación al alumnado del Grado en Matemáticas pasan a un segundo plano permitiendo que el aprendizaje se enfoque en el contenido propio de la asignatura que es el modelado y la resolución de problemas de optimización. Esta metodología aporta valor añadido al trabajo del alumnado pues el resultado se publica en entorno web para que pueda ser usado y consultado por cualquier persona interesada. Durante los cursos 14/15 y 15/16 se solicitaron y fueron apoyados sendos proyectos de Innovación Docente casi homónimos a la comunicación que se presenta a desarrollar en la asignatura de Programación Matemática dentro del Grado en Matemáticas en la Universidad de Cádiz. Si bien el nombre de la asignatura puede sugerir que se trata de una asignatura de programación informática no es así. La asignatura trata de la resolución de problemas de optimización de recursos y de planificación. En los cursos 16/17 y 17/18 se continuó con la dinámica introducida durante la ejecución de los citados proyectos. El resultado ha sido que se han desarrollado un total de siete aplicaciones interactivas en línea que se han publicado para uso público en <http://knuth.uca.es/shiny/> Estas aplicaciones han abordado problemas como:

- Caminos más cortos en grafos
- Localización de plantas sin capacidades
- Planificación de horarios

- Planificación de proyectos
- Ruta del viajante de comercio

Además, cada equipo participante ha desarrollado un curso, que también es accesible públicamente, en el que introducen el problema considerado y exponen las estrategias necesarias para su resolución. Además, dichos cursos incluyen ejemplos de problemas resueltos con las aplicaciones realizadas y manuales explicando el uso del software. Para que esto haya sido posible se ha desarrollado por parte de los docentes una aplicación “esqueleto” que gestiona toda la entrada y salida de datos. De esta forma se reduce mucho el esfuerzo necesario por parte del alumnado participante y se enfoca el aprendizaje en los aspectos más científicos de la programación. Todo este material se publica bajo licencia libre GNU-GPL o GNU-FDL según corresponda. La importancia que los problemas abordados tienen en nuestra vida diaria se pone de manifiesto fácilmente. Por ejemplo, cada vez que le indicamos a nuestro teléfono que nos indique una ruta para llegar a nuestro destino, el teléfono está resolviendo un problema de camino más corto. Este es uno de los problemas tratados. Otro ejemplo de gran importancia es el del problema del viajante de comercio. Todos los días las empresas transportistas tiene que resolver un problema del viajante de comercio para repartir todos los paquetes sin visitar dos veces a un mismo cliente. Este problema tiene una gran importancia económica y es objeto de múltiples publicaciones en revistas científicas de reconocido prestigio internacional. Los cursos están accesibles en <http://knuth.uca.es/moodle/cou>

Referencias/Bibliografía

- Alumnado de Programación Matemática and Manuel Muñoz Márquez. Cursos sobre las aplicaciones interactivas en línea para la resolución de problemas de optimización.
- Alumnado de Programación Matemática and Manuel Muñoz Márquez. Aplicaciones interactivas en línea para la resolución de problemas de optimización.
- Manuel Muñoz Márquez. Aplicación esqueleto para el desarrollo de aplicaciones interactivas en línea para la resolución de problemas de optimización.

C202. Explorando los datos públicos de España: un paquete para navegar la API de datos.gob.es

Jorge Cimentada^a, Jorge López^b

^aPompeu Fabra & Kernel Analytics, cimentadaj@gmail.com

^bKernel Analytics, jorge@loperez.com

Con el crecimiento exponencial de datos públicos en Europa, el gobierno Español ha publicado gran parte de sus datos de forma pública en una API unificada en la página <https://datos.gob.es/>. Hasta ahora, esta API agrega cerca de 19,000 ficheros de datos incluyendo temas como medioambiente, economía, demografía y educación entre otros muchos. Esta agregación de datos permite filtrar a nivel provincial, regional y a niveles más granulares como administraciones locales y universidades. Esta nueva fuente de datos presupone un salto importantísimo para la investigación cuantitativa española ya que permite analizar estos datos con el fin de descubrir patrones en la economía, en nuestra sociedad o en el sistema educativo. Estos datos también permiten evaluar la labor del gobierno de forma más cercana y exigir rendición de cuentas. Pero aún más importante, dado la riqueza y cantidad de los datos, permite unificar estos datos con encuestas u otros datos públicos más comunes para crear bases de datos más completas. Actualmente la API es fácil de manipular. El portal ofrece una documentación bastante detallada y permite descargar los datos en muchos formatos diferentes (incluyendo CSV, JSON, y XLS). No obstante, navegar la funcionalidad de una API puede ser complejo y esta no es una habilidad común o necesaria en el expertise de los científicos (con excepción de los científicos de datos y los informáticos). En esta ponencia queremos presentar un paquete de R que simplifica la extracción de datos de la API de datos.gob.es. El paquete OpendataES engloba la API de una manera que no requiere saber absolutamente nada de su funcionalidad. Actualmente el paquete tiene dos funcionalidades. En primer lugar, te permite navegar los datos disponibles en la API directamente desde R con un mínimo esfuerzo. Después de encontrar los datos de interés, se pasa la identificación de los datos y el paquete se encarga de descargarlos a R en los formatos más sencillos de leer. En la presentación haremos un live demo de lo que puede actualmente hacer el paquete y lo que no. Presentaremos los puntos a mejorar con miras a que los participantes del congreso se pueden convertir en contribuidores y avanzar el análisis de datos en España hacia mayor transparencia.

Creemos que este paquete puede abrir puertas para que la ciencia sea más abierta y reproducible. Los beneficios de este paquete no solo permite descargar datos al momento, si no que cualquier usuario con conocimientos de informática puede descargar datos que son publicados cada minuto (o segundos) y crear bases de datos históricas. Un ejemplo de esto es con los patrones de movimiento de bicicletas, que pudiera servir para estudiar nuevas formas de transporte sostenible. A pesar de esto, OpendataES está todavía en fase de maduración y actualmente solo se pueden leer aproximadamente un 70 % de los datos. Adicionalmente, la heterogeneidad de los datos presupone una limitación importante ya que los datos vienen de fuentes y formatos diferentes lo que implica que algunos datos son más difíciles de leer. Esta ponencia presentará el paquete por primera vez en público. El paquete ha estado en desarrollo por los últimos 3-4 meses y ya es momento de que vea la luz. Pensamos que OpendataES tiene mucho potencial para descargar datos, pero a la vez tiene muchas limitaciones que solo pueden ser resueltas con contribuciones de la comunidad de R y con más personas utilizándolo. Dicho esto, pensamos que el mejor lugar para dar a conocer el paquete es la Jornadas de Usuarios de R.

Paquetes de R que emplea: opendataes

C202. Explorando los datos públicos de España: un paquete para navegar la API de datos.gob.es

Jorge Cimentada^a, Jorge López^b

^aPompeu Fabra & Kernel Analytics, cimentadaj@gmail.com

^bKernel Analytics, jorge@loperez.com

Con el crecimiento exponencial de datos públicos en Europa, el gobierno Español ha publicado gran parte de sus datos de forma pública en una API unificada en la página <https://datos.gob.es/>. Hasta ahora, esta API agrega cerca de 19,000 ficheros de datos incluyendo temas como medioambiente, economía, demografía y educación entre otros muchos. Esta agregación de datos permite filtrar a nivel provincial, regional y a niveles más granulares como administraciones locales y universidades. Esta nueva fuente de datos presupone un salto importantísimo para la investigación cuantitativa española ya que permite analizar estos datos con el fin de descubrir patrones en la economía, en nuestra sociedad o en el sistema educativo. Estos datos también permiten evaluar la labor del gobierno de forma más cercana y exigir rendición de cuentas. Pero aún más importante, dado la riqueza y cantidad de los datos, permite unificar estos datos con encuestas u otros datos públicos más comunes para crear bases de datos más completas. Actualmente la API es fácil de manipular. El portal ofrece una documentación bastante detallada y permite descargar los datos en muchos formatos diferentes (incluyendo CSV, JSON, y XLS). No obstante, navegar la funcionalidad de una API puede ser complejo y esta no es una habilidad común o necesaria en el expertise de los científicos (con excepción de los científicos de datos y los informáticos). En esta ponencia queremos presentar un paquete de R que simplifica la extracción de datos de la API de datos.gob.es. El paquete OpendataES engloba la API de una manera que no requiere saber absolutamente nada de su funcionalidad. Actualmente el paquete tiene dos funcionalidades. En primer lugar, te permite navegar los datos disponibles en la API directamente desde R con un mínimo esfuerzo. Después de encontrar los datos de interés, se pasa a la identificación de los datos y el paquete se encarga de descargarlos a R en los formatos más sencillos de leer. En la presentación haremos un live demo de lo que puede actualmente hacer el paquete y lo que no. Presentaremos los puntos a mejorar con miras a que los participantes del congreso se pueden convertir en contribuidores y avanzar el análisis de datos en España hacia mayor transparencia.

Creemos que este paquete puede abrir puertas para que la ciencia sea más abierta y reproducible. Los beneficios de este paquete no solo permite descargar datos al momento, si no que cualquier usuario con conocimientos de informática puede descargar datos que son publicados cada minuto (o segundos) y crear bases de datos históricas. Un ejemplo de esto es con los patrones de movimiento de bicicletas, que pudiera servir para estudiar nuevas formas de transporte sostenible. A pesar de esto, OpendataES está todavía en fase de maduración y actualmente solo se pueden leer aproximadamente un 70 % de los datos. Adicionalmente, la heterogeneidad de los datos presupone una limitación importante ya que los datos vienen de fuentes y formatos diferentes lo que implica que algunos datos son más difíciles de leer. Esta ponencia presentará el paquete por primera vez en público. El paquete ha estado en desarrollo por los últimos 3-4 meses y ya es momento de que vea la luz. Pensamos que OpendataES tiene mucho potencial para descargar datos, pero a la vez tiene muchas limitaciones que solo pueden ser resueltas con contribuciones de la comunidad de R y con más personas utilizándolo. Dicho esto, pensamos que el mejor lugar para dar a conocer el paquete es la Jornadas de Usuarios de R.

Paquetes de R que emplea: opendataes

C203. Bringing Standards to Quality Control with R

Manuel Alfaro García^a, EmilioLópez Cano^b

^aUniv. de Castilla-La Mancha

^bUniv. de Castilla-La Mancha

Why standards?

Standards benefit both manufacturers and customers, from private an public sector. They bring benefits for all the society, by:

- Helping to optimize the management of the companies and the provision of services, taking costs down.
- Improving the interoperability between products and systems.
- Increasing acceptance of both products and services by establishing reference methods.
- Favoring communication and collaboration between organizations.
- Among others. . .

They provide the basis for mutual understanding, enable companies to comply with regulations, speed up the development of the products, facilitate interaction between business and establish a bridge between research and industry.

What is a standard?

As the International Organization for Standardization (ISO) says standards are the answer to the question,

What's the best way of doing this?

In a more formal way, CEN describes standards as,

“A standard is a technical document designed to be used as a rule, guideline or definition. It is a consensus-built, repeatable way of doing something.

Standards are created by bringing together all interested parties such as manufacturers, consumers, and regulators of a particular material, product, process or service. All parties benefit from standardization through increased product safety and quality as well as lower transaction costs and prices.”

Benefits of standards are countless, with a brief search on Google you can find up to 805.000.000 results explaining the advantage of implementing standards into your organization.

What does this have to do with R and Quality Control?

As we have said before benefits of standards are countless making this a reality into the industry. Most of the companies and organizations have implemented standards into their process management and most of their daily operations.

Industry and business base their operations and confidence in the use of standards. So any tool or process intended to be part of this world is practically forced to implement them.

ISO 9000 covers Quality Control, and define it as,

“A part of quality management focused on fulfilling quality requirements”

The 9000 series is designed to help companies to ensure they meet the requirements and the needs related to a service or product.

When we came into the statistical area of Quality Control is the **ISO Technical Committee 69, Statistical Methods** the one in charge of providing the standards. They have published 109 standards and into their subcommittees, they cover most of the aspects related to statistical methods.

- ISO/TC 69/SC 1 Terminology and symbols
- ISO/TC 69/SC 4 Applications of statistical methods in product and process management
- ISO/TC 69/SC 5 Acceptance sampling
- ISO/TC 69/SC 6 Measurement methods and results

4. RESÚMENES DE PONENCIAS

- ISO/TC 69/SC 7 Applications of statistical and related techniques for the implementation of Six Sigma
- ISO/TC 69/SC 8 Application of statistical and related methodology for new technology and product development

So standards are base to provide methods and guidelines to perform quality control. They also establish the base of the statistical methods and how they must be applied.

The industry is claiming for a standard way of applying statistical quality control and improvement into their process, by bringing both to R, standards and quality control, we can build the solution that the industry is waiting for.

RISSQ

R is becoming more and more popular, it is on the top positions of the programming languages rankings. It is growing up thanks to the big community that supports and contribute to build and maximize the functionality of the language.

So our plan is to build an Infrastructure for Statistical Standards on Quality to offer the industry a way of performing statistical quality control methods based on the standards with the software R.

With RISSQ we want to face some problems that arise with the actual software for SPC. The most important issue both with commercial software and open source tools is the **variety of the results** for the same data inputs, mostly small differences between platforms. But this can lead to future problems and difficulties may arise to meet the specifications and requirements of the products or services.

It is critical to have the same results no matter the software you use to perform the analysis. Standards provide guidelines to implement the different methods and datasets with inputs and expected results to check the implementation of these.

Another problem we want to face is **flexibility**, with an R infrastructure we allow users to adapt the software for their needs. That means the user can use the software as a final solution with a Shiny App and perform the analysis with no need of programming or knowledge about how the methods are implemented or what is running under the hood.

The user can also extend the library and use the base infrastructure we provide to build a new solution and perform more complex analysis and reports. This can be easily made with the R software, by just installing the packages and reading the documentation a third party developer can code a personalized solution for his problem.

R Implementation

For this we have built RISSQ following the standards, the methods have been implemented upon the guidelines the norms offers. And we have established a universal data structure following the standards to favor the communication between applications and different groups, implemented with S4 classes.

The functionality is implemented in the following packages.

- rissq: common structures and classes
- rissq.data: data sets and examples
- rissq.app: shiny App
- rissq.io: utilities for input and output tasks (import, export, etc.)

Also, the statistical methods are defined in different packages by their purpose.

- rissq.spc: Statistical Process Control (Control Charts, Capability)
- rissq.msa: Measurement Systems Analysis (Type I, Crossed, Nested, Unbalanced)
- rissq.as: Acceptance sampling

Similar to the philosophy of the tidyverse package family we have built the infrastructure trying to get:

- Common data structures across packages, in order to aim the interaction between projects.
- Solve complex problems by composing simple pieces, by dividing the functionality and the implementation of the different methods everything is more readable and easy to understand.
- Embrace FP, even if the data structures and the different methods are built upon the S4 model the users will still be able to use the package with functions.
- Write for humans, both the code and the documentation have been developed in the most simple way to help new users get into RISSQ.

Current work

Right now we are working on implementing a complete path for Measurement System Analysis (MSA), define de classes, methods, generics and launch the application to allow the user perform the analysis from the web interface.

Then check the methods against the standards and receive feedback from the users in order to see how it is the software performing with real environments and data.

Built the website and document the infrastructure.

After implementing a complete path the next step is to do the same with all the different methods.

C204. The efficiency of public lighting service in Spain

Pedro-José Martínez Córdoba^a; María-Dolores Guillamón López^b; Bernardino Benito López^c; Úrsula Faura Martínez^d

^aUniversidad de Murcia

^bUniversidad de Murcia

^cUniversidad de Murcia

^dUniversidad de Murcia

La eficiencia de los servicios públicos se ha convertido en una prioridad tras la reforma constitucional de septiembre de 2011. En particular, las Administraciones Locales están sujetas a cumplir el equilibrio presupuestario en sus cuentas, con estricta sujeción a la normativa de estabilidad presupuestaria y sostenibilidad financiera. Los Ayuntamientos son los encargados de prestar el servicio de alumbrado público, que es obligatorio para todos ellos y supone un importante porcentaje del presupuesto municipal, además de la peculiaridad de ser considerado un bien público puro. Con los datos de coste efectivo y las unidades físicas de referencia, analizamos la eficiencia de este servicio para los municipios españoles de más de 1.000 habitantes. Concretamente, utilizamos el Análisis de Datos Envoltentes (package r-DEA), una técnica matemática internacionalmente aceptada, y evaluamos la eficiencia de los municipios para los años 2014, 2015 y 2016. Seguidamente analizamos un grupo de variables de diferente naturaleza (geográficas, políticas, económicas y de gestión) para determinar el efecto que producen en la eficiencia (package truncreg). Los resultados de nuestra investigación muestran que la gestión privada produce ineficiencia para este servicio, a diferencia de la gestión pública que tiene un efecto positivo sobre la eficiencia. Respecto a las variables geográficas, muestran un efecto negativo, es decir, una mayor densidad de población y una mayor superficie urbanizada tendrá como consecuencia menor eficiencia del alumbrado público. Las horas de sol del término municipal afectan positivamente a la eficiencia, concretamente más horas de sol se traducen en más eficiencia del servicio. Respecto a las variables políticas, gobiernos conservadores y fragmentados serán menos eficientes que gobiernos progresistas y con mayoría en la corporación municipal. Finalmente, la variable económica de ingresos municipales presenta un efecto negativo; en este sentido, un Ayuntamiento con más ingresos será más ineficiente, ya que dispondrá de más presupuesto y en consecuencia la racionalidad del gasto no será estricta. Por otra parte, el índice turístico del municipio presenta un efecto positivo sobre la eficiencia del alumbrado público, es decir, los municipios turísticos serán más eficientes en este servicio.

Paquetes de R que emplea: r-DEA , productivity

C204. LAU2boundaries4spain: un paquete de R con contornos municipales para el periodo 2012-2018

Pedro J. Pérez^a , Francisco Goerlich^b

^aUniversitat de València, pedro.j.perez@uv.es

^bUniversitat de València e IVIE

LAU2boundaries4spain es un paquete de R cuyo objetivo es ofrecer los contornos de los términos municipales españoles ajustados a los municipios existentes a 1 de enero de cada año desde 2002, de forma que cuadren exactamente con el Padrón de dicho año. Creemos que dicha base de datos es útil por varias razones. El Instituto Geográfico Nacional (IGN) ofrece, a través del Centro de

Descargas del Centro Nacional de Información Geográfica, los recintos municipales –también los provinciales y autonómicos– dentro de la Base de Datos de Líneas de Límite Municipal en la Información Geográfica de Referencia, sin embargo estas bases de datos son de actualización continua, de forma que lo que se dispone son los municipios “en el momento actual”, y no existe un histórico que puede ser de utilidad por muchas razones. Así pues, hasta donde nosotros conocemos, no se dispone, por ejemplo, de una capa vectorial de los polígonos municipales del año 2006, o de la de los municipios correspondientes al censo de 2001. Esta información es necesaria si queremos combinar la información geográfica con la información estadística municipal histórica procedente, por ejemplo, del Instituto Nacional de Estadística (INE). Este repositorio pretende cubrir esa laguna que, de momento, no ha sido satisfecha por las instituciones oficiales. El repositorio está disponible como un package de R alojado en Github: <https://github.com/perezp44/LAU2boundaries4spain>. La información ofrecida ha ido elaborándose durante bastante tiempo a partir de diversos trabajos (Goerlich, Mas, Azagra y Chorén 2006, 2007; Goerlich, Ruiz, Chorén y Albert 2015; Reig, Goerlich y Cantarino 2016). Información más detallada del proceso seguido para la construcción de los diferentes ficheros de lindes puede encontrarse en una de las vignettes del package (<<https://htmlpreview.github.io/?https://github.com/perezp44/LAU2boundaries4spain/blob/master/inst/doc/detailed-info-lau2boundaries4spain.html>>)

¿Paquetes de R que emplea: sf, tidyverse

Referencias/Bibliografía

- J.A. Ros, P.C. Rodríguez, F.J.G. Gisbert, M. Mas, and BBVA (Firm). Fundación. La localización de la población española sobre el territorio: un siglo de cambios : un estudio basado en series homogéneas (1900-2001). Elibro Catedra. Fundación BBVA, 2006.
- F.J.G. Gisbert, F.R. Rodríguez, P.C. Rodríguez, and C.A. Pérez. Cambios en la estructura y localización de la población: una visión de largo plazo (1842-2011). Fundación BBVA, 2015.
- E.R. Martínez, F.J.G. Gisbert, and I.C. Martí. Delimitación de áreas rurales y urbanas a nivel local: Demografía, coberturas del suelo y accesibilidad. Economía y sociedad. Fundación BBVA, 2016.

Ponencia Invitada II

Text Mining Using Tidy Data Principles

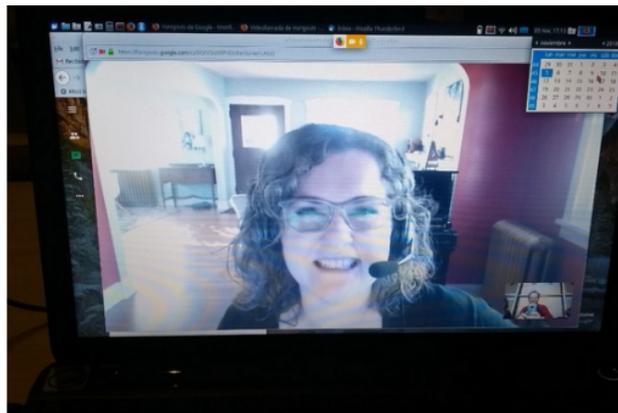
Julia Silge^a

^aData scientist at Stack Overflow, julia.silge@gmail.com



JULIA SILGE

Text data is increasingly important in many domains, and tidy data principles and tidy tools can make text mining easier and more effective. In this talk, Julia will demonstrate how we can manipulate, summarize, and visualize the characteristics of text using these methods and R packages from the tidy tool ecosystem. These tools are highly effective for many analytical questions and allow analysts to integrate natural language processing into effective workflows already in wide use. We will explore how to implement approaches such as sentiment analysis of texts, measuring tf-idf, and finding word vectors.



5

Talleres

Cuadro 5.1: **Talleres del Jueves** (id es un enlace a los materiales)

id	Título	Autorías (entidad)
J1	Creación de mapas interactivos con R y Leaflet	Antonio Sánchez Chinchón (Telefónica)
J2	Evaluación de las medidas de capacidad predictiva obtenidas con técnicas de remuestreo mediante un test de significación	Jesús Herranz (GEICAM)
J3	Redes Bayesianas con R	Bojan Mihaljević (Univ. Politécnica de Madrid)

Cuadro 5.2: **Talleres del Viernes** (id es un enlace a los materiales)

id	Título	Autorías (entidad)
V1	Text Mining & NLP con R: Desde el análisis teóricos de textos en Esperanto hasta las Aplicaciones Empresariales con shiny	Francisco Jesús Rodríguez Aragón (Teradata-Think Big Analytics)
V2	Análisis de correspondencias con R: aplicación a (micro) datos de encuestas	Emilio López Cano (Univ. de Castilla-La Mancha)

Taller J1: Creación de mapas interactivos con R y Leaflet

Antonio Sánchez Chinchón^a

^aTelefónica, antonio.sanchezchinchon@telefonica.com

Con R se pueden hacer mapas estáticos pero a veces es necesario interactuar con ellos para entender mejor qué nos quieren decir. Leaflet es una biblioteca muy popular de mapas interactivos escrita en JavaScript que se integra perfectamente con R y te permitirá hacer mapas muy atractivos y profesionales con poco esfuerzo. En este taller empezaremos creando mapas muy simples, en los que colocaremos unos cuantos puntos y poco a poco iremos complicándolos hasta llegar a trabajar con polígonos que delimitan fronteras. No hace falta conocer Leaflet y tan sólo es necesario un conocimiento básico de R y RStudio para aprovecharlo. Aprenderemos a guardarlos, a cambiar el mapa base y a incluir popups. Además, aprenderemos a obtener las coordenadas geográficas de cualquier lugar del mundo. Todo esto te ayudará a crear mapas para visualizar datos desde una perspectiva geográfica. Después de este taller podrás aprovecharte del hecho de que todo ocurre en algún lugar del mundo.

Taller J2: Evaluación de las medidas de capacidad predictiva obtenidas con técnicas de remuestreo mediante un test de significación

Jesús Herranz^a

^aGEICAM, jherranzvalera@yahoo.es

Las técnicas de remuestreo, como la validación cruzada (CV) o el bootstrapping, se han convertido en las técnicas estándar para la obtención de medidas de capacidad predictiva de los modelos construidos con cualquier técnica de machine learning o de estadística predictiva. Este taller pretende ir un paso más allá, y plantear si esas medidas de capacidad predictiva, obtenidas por remuestreo, son diferentes de las que se obtendrían por un modelo que predice aleatoriamente. Por ejemplo, en un problema de clasificación, donde el área bajo la curva ROC (AUC) es una medida de capacidad predictiva muy usada, se trataría de plantear si el AUC obtenido por remuestreo es diferente de 0.5. En un problema de regresión, donde el error cuadrático medido (MSE) es considerado como la medida de capacidad predictiva más común, se trataría de plantear si el MSE obtenido por remuestreo es diferente del obtenido por un modelo que predice aleatoriamente, es decir, de un error cuadrático medio basal.

Por tanto, se está proponiendo un test de significación para conocer si nuestro modelo predictivo predice o no mejor que uno que predice de forma aleatoria, o dicho de otra forma, si la medida obtenida en nuestro modelo (AUC, MSE o cualquier otra) podría haber sido obtenida con un modelo sin ninguna capacidad predictiva.

Por la forma en la que estas medidas de capacidad predictiva son construidas, si se quiere obtener un test para evaluar si realmente el modelo predice, es necesario hacerlo mediante un test de permutaciones. Este taller muestra la metodología general que hay que seguir para construir esta clase de test en cualquier situación, es decir, con cualquier técnica de construcción de modelos predictivos y con cualquier medida de capacidad predictiva, y lo más importante, cómo programar este proceso complejo en R de una forma ordenada. Además, dada la naturaleza del procedimiento, se explicará también una implementación alternativa con técnicas de paralelización.

El taller es fundamentalmente un taller de metodología estadística, pero es importante ser consciente que estas metodologías tan complejas solo es posible implementarlas en lenguajes de programación tan flexibles como es R, donde es fácil integrar todas las fases de este proceso, incluidas las técnicas de machine learning más complejas.

En el taller se repasará brevemente el proceso completo para construir modelos predictivos, que incluye las siguientes fases: selección de variables, optimización de parámetros y construcción del modelo con los parámetros optimizados. Se repasará también la validación cruzada como técnica de remuestreo, y cómo programarla en R. Finalmente, se explicará cómo diseñar un test de permutaciones en esta situación para generar la distribución de la medida de capacidad predictiva de nuestro modelo bajo la hipótesis nula de que nuestro modelo no predice (AUC=0.5, por ejemplo); para finalmente obtener un p-valor asociado a la medida de capacidad predictiva obtenida por remuestreo.

Finalmente, se mostrará la metodología general de cómo reprogramar con paralelización procesos de este tipo usando el paquete parallel, lo cual puede ser también de interés para asistentes que necesiten paralelizar otros procesos.

Todo el taller se desarrollará explicando simultáneamente los conceptos teóricos y la programación en R, basada en un ejemplo concreto.

Por último comentaros, que he participado en otras Jornadas de R con dos talleres: *¿Análisis de Supervivencia con datos de Alta Dimensionalidad?* presentado en las VII Jornadas (Salamanca, 2015), y *“Modelos Predictivos con el paquete caret”* en las VI Jornadas (Santiago de Compostela, 2014)

Taller J3: Redes Bayesianas con R

Bojan Mihaljević^a

^aUniv. Politécnica de Madrid, bmihaljevic@fi.upm.es

Resumen: Las redes Bayesianas pueden representar distribuciones de probabilidad sobre muchas variables aleatorias al aprovechar independencias condicionales. Los objetivos del tutorial son:

- Introducir los fundamentos de las redes Bayesianas e ilustrarlos con ejemplos prácticos.
- Presentar las funcionalidades que ofrece R para el aprendizaje a partir de datos y la inferencia con redes Bayesianas, explorando importantes aspectos prácticos.

Utilizaremos los paquetes `bnlearn` y `gRain`.

Esquema

- Introducción teórica: Conceptos básicos. Propiedades de Markov. Distribuciones condicionadas discretas y Gausianas.
- Aprendizaje de estructura: algoritmos constraint-based y score-based.
- Aprendizaje de parámetros.
- Inferencia: exacta `gRain` y aproximada con `bnlearn`.

Taller V1: Text Mining & NLP con R: Desde el análisis teóricos de textos en Esperanto hasta las Aplicaciones Empresariales con shiny

Francisco Jesús Rodríguez Aragón^a

^aTeradata-Think Big Analytics, fjroar@hotmail.com

En el taller se pretende hacer en primer lugar una introducción a través de un primer recorrido por las técnicas más usuales del Text Mining y de las técnicas NLP (Natural Language Processing).

Posteriormente y como primeras aplicaciones se analizan ejemplos enmarcados en el análisis de textos largos donde se trata de hacer comparativas de libros ya digitalizados en distintas lenguas con respecto al Esperanto donde se analiza que rasgos en común y distintos existen entre lenguajes naturales como el inglés y lenguajes artificiales.

Tras tener unos conceptos básicos se tratará de realizar aplicaciones reales con textos cortos anonimizados y que tratarán de ser clasificados de modo automático y coherente a través de modelos matemáticos

Finalmente y en base al ejemplo anterior se mostrará como puede crearse un aplicativo shiny básico que permite la paquetización de lo realizado y su puesta de producción.

Así pues, el objetivo del taller además de hacer un recorrido por las técnicas habituales del text mining, también permitirá la práctica y el recorrido por las distintas fases por las que un proyecto de estas características debe pasar, como es la fase del “puro discovery” hasta su paquetización en una app final que permita articular los algoritmos desarrollados.

Taller V2: Análisis de correspondencias con R: aplicación a (micro) datos de encuestas

Emilio López Cano^a

^aUniv. de Castilla-La Mancha, emilio.lcano@uclm.es

Objetivo: El objetivo general de este taller es doble: aprender a realizar análisis de correspondencias con R y practicar el tratamiento de datos de encuestas disponibles públicamente.

Justificación: Los resultados de los estudios realizados por diversos organismos públicos como el Instituto Nacional de Estadística (INE), el Centro de Investigaciones Sociológicas (CIS), EUROSTAT, y muchos otros (incluidos organismos regionales), se han publicado tradicionalmente en forma resumida o con las principales conclusiones. Tiene especial interés para investigadores de todo tipo el analizar los datos *originales* de las encuestas, esto es, los microdatos. Las tendencias actuales en transparencia y *Open Data* están haciendo que cada vez más organismos publiquen estos microdatos de forma que puedan ser utilizados por terceros para

realizar sus estudios. Esta es una fuente de información muy rica que se puede explotar en investigación de mercados, de salud pública, economía, sociología, etc. Por otra parte, los datos que contienen las encuestas son principalmente cualitativos, por lo que el análisis de correspondencias, que precisamente pretende identificar patrones, tendencias y relaciones entre atributos.

Audiencia: El taller es adecuado para cualquier analista que trate datos cualitativos de cualquier tipo. Es especialmente útil para quienes trabajen con datos de encuestas, sea cual sea el ámbito (socioeconómico, sanitario, psicología, marketing, etc.), y en general para cualquier persona interesada en R y/o el análisis de datos.

Contenido: En este taller vamos a trabajar con la última Encuesta europea de salud en España (EESE) que elabora el Instituto Nacional de Estadística. No obstante, las técnicas son de aplicación a cualquier otra encuesta (EPA, EGM, ...). La diferencia va a estar generalmente en la forma de importar los datos en forma de tabla para poder trabajar con ellos en R.

En primer lugar, se revisarán los documentos metodológicos que acompañan a las publicaciones del INE. Así, la información relevante de cada estudio contendrá:

1. Informes de resultados
2. Microdatos
3. Documentos metodológicos
 - Los cuestionarios
 - Metodología general
 - Diseño de registro

La metodología general suele ser un documento bastante extenso, pero con las primeras páginas, los cuestionarios, y el diseño de registro, debemos ser capaces de averiguar el formato de los microdatos para importarlos y tratarlos con R. Algunos microdatos se pueden tratar fácilmente con el paquete `MicroDatosEs`, aunque en este taller se tratan datos más *crudos* de otras encuestas.

Una vez revisados los documentos relevantes y los ficheros de datos, procederemos a la importación de los datos al entorno de trabajo de R desde RStudio. Los datos de la EESE están disponibles en formato texto plano de ancho fijo, que se pueden importar de forma muy eficiente con ayuda del diseño de registro. Pero también están en formato Excel, cuya importación es trivial con el paquete `readxl` por lo que se usará este formato para el taller.

A continuación, se realizarán para algunas variables seleccionadas de esta encuesta análisis de correspondencias simples y múltiples, principalmente utilizando el paquete `FactoMineR` y también `factoextra` para algunas visualizaciones.

Finalmente, se darán algunas pinceladas de temas avanzados de R para que los asistentes exploren en el futuro y saquen aún más provecho de las técnicas.

Material y Requisitos: El material del taller estará en un repositorio github en forma de proyecto de RStudio descargable con función `usethis::use_course`. Recomendado tener paquetes a usar.

Sobre el ponente: Emilio López Cano, entusiasta y activista de R. Profesor en la Universidad de Castilla-La Mancha. Autor de dos libros en la serie Use R! de Springer y creador del paquete `SixSigma`. Formador en AEC

6

Pósters

Póster 1. Intelligent Climate. A Machine Learning Solution built on SAP Leonardo & R for HVAC Energy and Comfort optimization

Manuel Jesús Rodríguez Delgado, Antonio Espejo Écija, Christian Fernández Calvillo, Manuel Gutiérrez Plaza, Ismael Moreno Mechbal, Jesús Figueres Cañadas, Jonathan Fuentes Gutiérrez, Alejandro Alija^a, Mario E. López López

^aAccenture, alejandro.alijs@accenture.com

Most of air-conditioning and heating systems (HVAC) follow a very traditional approach when ruling the environmental management programs, based on hardcoded rules (RBC Rule-Based Controllers) that command the whole system in an inflexible fashion. In our project, we have developed an intelligent climate solution. It involves machine learning models powered by R engine and integrated with Extended Application Services (XS); real-time data ingestion through the Internet of Things 4.0 service at SAP Cloud Platform (SCP); SAP HANA data model and storage; end-user interface built in Fiori framework with MQTT (communication protocol) integration to remotely send commands back to the indoor units, and hardware integration with BACnet industrial protocol.



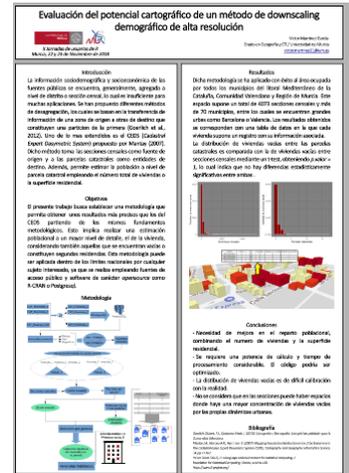
Póster 2. Evaluación del potencial cartográfico de un método de downscaling demográfico de alta resolución

Víctor Martínez García^a

^aUniversidad de Murcia, victor.martinez11@um.es

Autoría:

La información sociodemográfica y socioeconómica de las fuentes públicas (EUROSTAT, INE, Organismos de estadística regionales, etc) se encuentra, generalmente, agregada a nivel de distrito o sección censal. Este es insuficiente para muchas aplicaciones que requieren un mayor nivel de precisión. En la última década se han propuesto diferentes métodos de desagregación, los cuales se basan en la transferencia de información de una zona de origen a otras de destino de menor extensión, las cuales intersecan espacialmente con la primera. Uno de estos es el CEDS (Cadastral Expert Dasymetric System), muy extendido por la accesibilidad de sus datos (censo y catastro) y la calidad de sus resultados. Dicho método emplea las secciones censales como fuente de origen y las parcelas catastrales como entidades de destino. Para el reparto de la población considera que la proporción de individuos residente en la parcela con respecto al total de la sección es igual a la del número de viviendas o de la superficie residencial de cada parcela catastral con respecto al total de las parcelas contenidas en una misma sección censal. Es decir, que permite estimar la población a nivel de parcela catastral empleando dos variables distintas. El presente trabajo parte de los mismos fundamentos metodológicos y aumenta la precisión de los resultados obtenidos, realizando la estimación poblacional a un mayor nivel de detalle, el de la vivienda, y considerando aquellas que se encuentran vacías dentro de cada sección censal. El objetivo es establecer una metodología libremente aplicable dentro de los límites por cualquier sujeto interesado mediante el empleo de fuentes de acceso público y software de carácter opensource como R-CRAN o Postgresql.



Póster 3. Relación entre el perfil proteómico y el intervalo de la muerte en restos óseos humanos

Gemma Prieto Bonete^a, María Dolores Pérez Cárcelos^b, Antonio Maurandi López^c, Aurelio Luna Maldonado^d.

^aUniversidad de Murcia, gemma.prieto@um.es

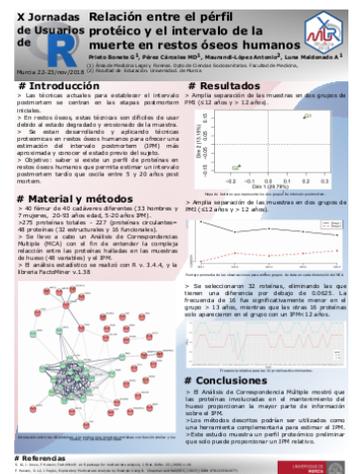
^bUniversidad de Murcia, mdperez@um.es

^cUniversidad de Murcia, amaurandi@um.es

^dUniversidad de Murcia

Las técnicas actuales para establecer el intervalo postmortem se centran en las etapas postmortem iniciales. En restos óseos, estas técnicas son difíciles de usar debido al estado degradado y erosionado de la muestra. Se están desarrollando y aplicando técnicas proteómicas en restos óseos humanos para ofrecer una estimación del intervalo postmortem (IPM) más aproximada y conocer el estado previo del sujeto. Objetivo: saber si existe un perfil de proteínas en restos óseos humanos que permita estimar un intervalo postmortem tardío que oscila entre 5 y 20 años post mortem.

- N. Procopio, A.T. Chamberlain, M. Buckley, Intra-and interskeletal proteome variations in fresh and buried bones, *J. Proteome Res.* 16 (2017) 2016-2029.
- N. Procopio, A. Williams, A.T. Chamberlain, M. Buckley, Forensic proteomics for the evaluation of the post-mortem decay in bones, *J. Proteom.* 177 (2018) 21-30.
- N. Procopio, A.T. Chamberlain, M. Buckley, Exploring biological and geological age-related changes through variations in intra-and inter-tooth proteomes of ancient dentine, *J. Proteome Res.* 17 (2018) 1000-1013.
- O. Unluturk, M.F. Sahin, The effect of remaining in water in estimation of time since N. Procopio, A.T. Chamberlain, M. Buckley, Intra-and interskeletal proteome variations in fresh and buried bones, *J. Proteome Res.* 16 (2017) 2016-2029.
- S. Pittner, B. Ehrenfellner, A. Zissler, V. Racher, W. Trutschnig, A.C. Bathke, F.C. Monticelli, First application of a protein-based approach for time since death estimation, *Int. J. Legal Med.* 131 (2017) 479-483.
- M. Kaliszán, R. Hauser, G. Kernbach-Wighton, Estimation of the time of death based on the assessment of post mortem processes with emphasis on body cooling, *Legal Med.* 11 (2009) 111-117.
- K. Makrodimitris, D.L. Masica, E.T. Kim, J.J. Gray, Structure prediction of protein;solid surface interactions reveals a molecular recognition motif of statherin for hydroxyapatite, *J. Am. Chem. Soc.* 129 (2007) 13713-13722.
- R.C. Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2014 URL <https://www.R-project.org/>.
- S. Lê, J. Josse, F. Husson, FactoMineR: an R package for multivariate analysis, *J. Stat. Softw.* 25 (2008) 1-18.
- F. Husson, S. Lê, J. Pagès, Exploratory Multivariate Analysis by Example Using R. Chapman and Hall/CRC, (2017) (ISBN 9781315301877).
- tributions in electrospray ionization mass spectrometry, *J. Mass Spectrom.* 38 (2003) 11-15.
- M.J. Gorbunoff, The interaction of proteins with hydroxyapatite: I. Role of protein charge and structure, *Anal. Biochem.* 136 (1984) 425-432.
- M.J. Gorbunoff, The interaction of proteins with hydroxyapatite: II. Role of acidic and basic groups, *Anal. Biochem.* 136 (1984) 433-439.



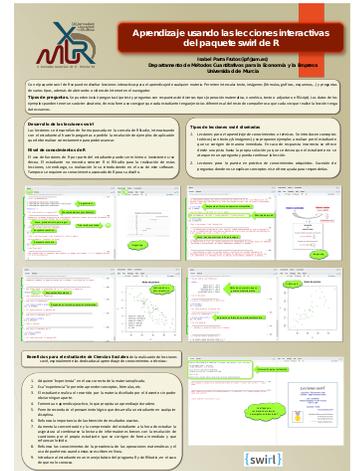
Póster 4. Aprendizaje usando las lecciones interactivas del paquete swirl de R

Isabel Parra Frutos^a

^aUniversidad de Murcia, ipf@um.es

En este curso 2018-2019 se han puesto en práctica, por primera vez, las lecciones swirl para los estudiantes de *Estadística básica para la Economía* de primero del grado de Economía de la Universidad de Murcia. Con el paquete swirl de R se pueden diseñar lecciones interactivas para el aprendizaje de cualquier tipo de materia. Permite intercalar texto, imágenes (con fórmulas, gráficos, esquemas,...) y preguntas de varios tipos. Se pueden incluir preguntas cuya respuesta puede venir en forma de selección entre varias opciones (cuestiones test), en forma de operación matemática, en forma numérica y en forma de texto. También hay preguntas para abrir una página web o para adjuntar como respuesta un RScript.

Las lecciones se desarrollan de forma pausada en la consola de RStudio, interactuando con el estudiante al hacerle preguntas o pedirle la resolución de ejemplos de aplicación que debe realizar correctamente para poder avanzar en la lección. Esta metodología permite al estudiante poner en práctica los conocimientos adquiridos (en la propia lección y/o previamente desarrollados en clase) obteniendo un feedback inmediato. Aunque se desarrollan en la consola de R, el uso de funciones de R por parte del estudiante puede ser mínimo o inexistente si se desea. El estudiante no necesita conocer R ni RStudio para la realización de estas lecciones, sin embargo, su realización le va introduciendo en el uso de este software. Tampoco se requiere un conocimiento avanzado de R para su diseño.



Las lecciones swirl están originalmente diseñadas para aprender R (“Learn R, in R”, <https://swirlstats.com/>). Sin embargo, este tipo de lecciones se pueden usar para el aprendizaje de cualquier materia (cuantitativa o no cuantitativa). Nosotros las hemos utilizado para el aprendizaje de Estadística Descriptiva. Hay que resaltar que hemos buscado el aprendizaje y comprensión de la Estadística Descriptiva pero no el análisis descriptivo de datos con R, que sería algo diferente. Dado que permiten el uso de cuestiones tipo test, cuestiones con respuesta numérica y alfanumérica (texto), la realización de lecciones swirl proporciona beneficios en cualquier disciplina, no solo en las cuantitativas. En el caso de materias cuantitativas se puede, además, combinar problemas de cálculo con cualquier otro tipo de preguntas para comprobar que el estudiante ha asimilado lo explicado o para, simplemente, practicarlos. Los datos de los ejemplos o ejercicios pueden tener un carácter aleatorio, de esta forma se consigue que cada estudiante tenga ejercicios diferentes al del resto de compañeros o que cada vez que realice la lección tenga datos nuevos.

Para controlar la realización de la lección por parte de los estudiantes hemos añadido al final una pregunta para que proporcionen su DNI. A partir de él, se calcula y se proporciona un código que debe entregar y que es único para cada lección y estudiante. También existe la opción de establecer una contraseña de inicio y de entregar el log de la lección para recuperar la información de la lección realizada por el estudiante (intentos realizados en las preguntas planteadas, etc.) mediante Google Forms.

Hemos elaborado dos tipos de lecciones swirl:

- *Lecciones para el aprendizaje de conocimientos o técnicas.* Se introducen conceptos teóricos (con texto y/o imágenes) y se proponen ejemplos a realizar por el estudiante que se corrigen de manera inmediata. En caso de respuesta incorrecta se ofrece desde una pista hasta la propia solución ya que se desea que el estudiante no se atasque en una pregunta y pueda continuar la lección.
- *Lecciones para la puesta en práctica de conocimientos adquiridos.* Consisten en una sucesión de preguntas donde no se explican conceptos ni se ofrece ayuda para responderlas.

Algunas características adicionales de las lecciones swirl son las siguientes:

- Intercalan, por defecto, mensajes de ánimo en inglés (se pueden eliminar o cambiar el idioma).
- Informan continuamente del porcentaje de lección completada.
- Permiten salir temporalmente de la lección para hacer cálculos y luego volver a la pregunta donde se dejó.
- Guardan el progreso de una lección si se abandona antes de terminarla.

La realización de lecciones swirl, especialmente las dedicadas al aprendizaje de conocimientos o técnicas, tiene una serie de beneficios positivos para el estudiante de Ciencias Sociales:

- Adquiere “experiencia” en el uso correcto de la materia explicada.
- Esa “experiencia” que adquiere le permite aprender conceptos, fórmulas, etc.
- El estudiante realiza el recorrido por la materia diseñado por el docente sin poder obviar ninguna parte.
- Fomenta un aprendizaje activo, lo que propicia un aprendizaje duradero.
- Permite reconducir el pensamiento lógico que desarrolla un estudiante en cualquier disciplina.
- Refuerza la importancia de la obtención de resultados exactos, sin errores.
- Aumenta la concentración y la comprensión del estudiante a la hora de estudiar la asignatura al combinarse la lectura de informaciones breves con la resolución de cuestiones por el propio estudiante que se corrigen de forma inmediata y que refuerzan lo leído.
- Refuerza los conocimientos de la prevalencia de las operaciones matemáticas y el uso de paréntesis cuando estas se escriben en línea.
- Introduce al estudiante en un manejo básico del programa R y de RStudio, en el caso de que no lo conozca.

Póster 5. Distribuciones tipo fase en memorias RRAM con R

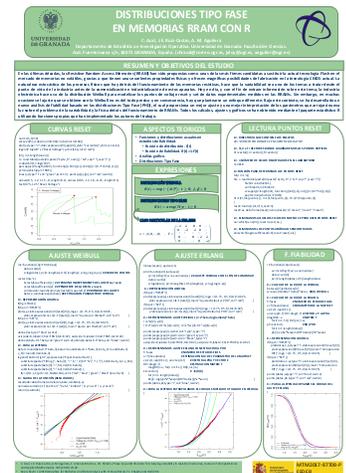
Christian Jose Acal González^a, Juan Eloy Ruiz Castro^b y Ana Maria Aguilera Del Pino^c

^aDpto. de Estadística e Investigación Operativa. Universidad de Granada. Facultad de Ciencias., chracal@correo.ugr.es

^bDpto. de Estadística e Investigación Operativa. Universidad de Granada. Facultad de Ciencias

^cDpto. de Estadística e Investigación Operativa. Universidad de Granada. Facultad de Ciencias

En las últimas décadas, las Resistive Random Access Memories (RRAM) han sido propuestas como unas de las más firmes candidatas a sustituir la actual tecnología Flash en el mercado de memorias no volátiles, gracias a que tienen unas excelentes propiedades físicas y ofrecen magníficas posibilidades de fabricación en la tecnología CMOS actual. La naturaleza estocástica de los procesos físicos que hay detrás del funcionamiento de las memorias resistivas, hace que la variabilidad sea uno de los temas a tratar desde el punto de vista de la industria antes de la comercialización e industrialización de estos aparatos. Hoy en día, y con el fin de extraer información sobre este tema, la industria electrónica hace uso de la distribución Weibull para modelizar los puntos de voltaje reset y set de datos experimentales medidos en las RRAMs. Sin embargo, en muchas ocasiones el ajuste que se obtiene con la Weibull no es del todo preciso y en consecuencia, hay que plantearse un enfoque diferente. Bajo este contexto, se ha desarrollado un nuevo análisis de fiabilidad basado en las distribuciones Tipo Fase, el cual proporciona un mejor ajuste y una mejor interpretación de los parámetros que arrojará nueva luz sobre el problema de la variabilidad y la física detrás del funcionamiento de RRAMs. Todos los cálculos, ajustes y gráficos se han obtenido mediante el paquete estadístico R utilizando funciones propias que han implementado los autores del trabajo. Este trabajo ha sido respaldado por el Ministerio de Economía y Competitividad, España, bajo la subvención MTM2017-87708-P y por el Fondo Europeo de Desarrollo Regional (FEDER)



Póster 6. Uso de aplicaciones interactivas Shiny para mejorar el rendimiento en las asignaturas de estadística

Juan José Pérez Castejón^a; Fuensanta Arnaldos García^b; M^a Teresa Díaz Delfa^c; Úrsula Faura Martínez^d; Lourdes Molera Peris^e

^aUniversidad de Murcia, jjpc@um.es

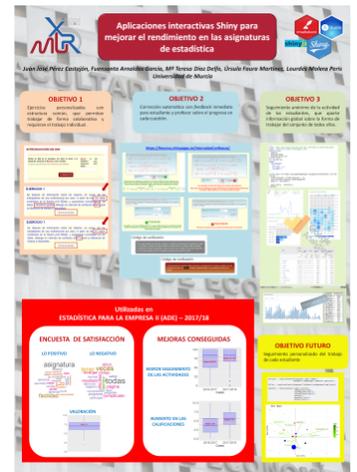
^bUniversidad de Murcia, arnaldos@um.es

^cUniversidad de Murcia, mdd@um.es

^dUniversidad de Murcia, faura@um.es

^eUniversidad de Murcia, lmolera@um.es

Con la adopción del Espacio Europeo de Educación Superior, el estudiante se ha convertido en el eje fundamental del proceso educativo, y fomentar el desarrollo de habilidades de autoaprendizaje es ahora una responsabilidad del docente. En este contexto, plantear actividades interactivas puede ser muy útil para motivar a los estudiantes y alentar en ellos una actitud activa y autónoma en la búsqueda del conocimiento. El software estadístico R es una herramienta libre, muy utilizada en la investigación y también en la docencia, con una gran versatilidad para el diseño de estrategias interactivas que favorezcan el aprendizaje autónomo. Una de sus mayores fortalezas es la gran cantidad de paquetes que posee y que de manera altruista aporta la comunidad de usuarios. Uno de estos paquetes, Shiny, permite la creación de aplicaciones web interactivas, que pueden emplearse tanto para apoyar la comprensión de conceptos y técnicas de estadística como para favorecer la autoevaluación. Con esta herramienta nuestro grupo de trabajo ha diseñado ejercicios que facilitan el aprendizaje de conocimientos o técnicas estadísticas usualmente complicadas para los estudiantes. Se trata de ejercicios personalizados a partir del DNI del estudiante, pero con una estructura común que permite trabajar de forma colaborativa. Los ejercicios se han planteado, además, para proporcionar un feedback inmediato al alumno sobre su progreso, incluyendo comentarios a sus respuestas, tanto cuando son correctas (con explicaciones u observaciones sobre posibles alternativas) como cuando son erróneas (con indicaciones para redirigir el aprendizaje). Una vez completado con éxito el ejercicio, el alumno recibe un código de verificación personal, que facilita al docente el desarrollo de una evaluación continua rigurosa, incluso en grupos numerosos, al corregirse de manera automática.



Estas aplicaciones se han utilizado en diferentes asignaturas de estadística en la Facultad de Economía y Empresa de la Universidad de Murcia. Concretamente, se están empleando de forma intensiva en la asignatura de Estadística para la Empresa II, del segundo curso del Grado en Administración y Dirección de Empresas, dedicada al estudio de la estadística inferencial. En esta asignatura se han diseñado ejercicios interactivos que acompañan a cada una de las prácticas. Estos ejercicios han sido testados en el primer cuatrimestre del curso 2017/18 y se dispone de información sobre la opinión de los alumnos que los han usado, así como de sus resultados académicos. Durante este curso se están utilizando de nuevo, con funcionalidades adicionales que permiten el seguimiento pormenorizado del trabajo individual de cada alumno, más allá del seguimiento general de la actividad global hecho a través de herramientas de análisis externas Google Analytics en el curso pasado. Además, se están traduciendo todas ellas al inglés para utilizarlas en el grupo bilingüe que se imparte en el Grado en Administración y Dirección de Empresas. La experiencia, tanto en el curso pasado como en el presente, ha resultado muy positiva para profesorado y alumnado. Los docentes hemos percibido una mejora en el manejo autónomo de R por parte de los estudiantes, así como una mayor comprensión de los conceptos y métodos estadísticos. Mediante una encuesta anónima al final del cuatrimestre del curso 2017/18, los alumnos han mostrado su opinión sobre estas aplicaciones interactivas indicando, de forma mayoritaria, que les han ayudado a entender mejor las nociones y técnicas estadísticas estudiadas. Además, destacan la facilidad de uso y el hecho de haber podido practicar con más ejercicios. Junto a la percepción cualitativa positiva de las aplicaciones, hemos comprobado de forma cuantitativa que, con su introducción, los alumnos realizan un mayor porcentaje de trabajos en su carpeta de actividades de evaluación continua, y esto se traduce, posteriormente, en mejores calificaciones finales.

Póster 7. Docencia con R en el departamento de Métodos Cuantitativos para la Economía y la Empresa

Lourdes Molera Peris^a, Juan José Pérez Castejón^b, Fuensanta Arnaldos García^c, M^a Teresa Díaz Delfa^d, Úrsula Faura Martínez^e, Isabel Parra Frutos^f, Susana Álvarez Díez^g, Francisco Martínez Sánchez^h, Pepe García Clavelⁱ

^aUniversidad de Murcia, lmolera@um.es

^bUniversidad de Murcia, ijpc@um.es

^cUniversidad de Murcia, arnaldos@um.es

^dUniversidad de Murcia, mdd@um.es

^eUniversidad de Murcia, faura@um.es

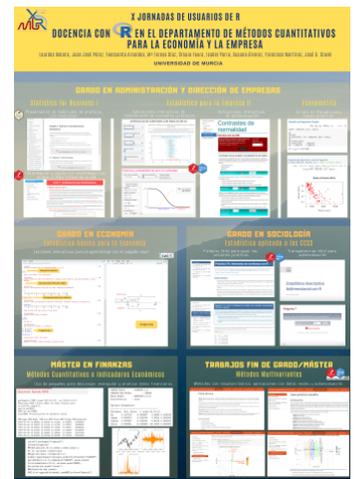
^fUniversidad de Murcia, ipf@um.es

^gUniversidad de Murcia, salvarez@um.es

^hUniversidad de Murcia, fms@um.es

ⁱUniversidad de Murcia

Tradicionalmente, en las asignaturas de Estadística y Econometría de la Facultad de Economía y Empresa de la Universidad de Murcia se han venido utilizando la hoja de cálculo Excel y los paquetes estadísticos SPSS y EViews. A pesar de la potencia y la versatilidad de R para el análisis de datos, su implementación en las clases prácticas de estas asignaturas no se ha establecido de forma definitiva hasta el uso generalizado de la interfaz RStudio, con todas las facilidades que incorpora. Las ventajas de R y RStudio son muchas, en especial su carácter gratuito y colaborativo, con infinidad de paquetes que pueden ser utilizados libremente por cualquier usuario, así como su versatilidad, puesto que favorecen el desarrollo del proceso de análisis de datos completo, desde la manipulación preliminar a la presentación de resultados, de forma reproducible. Además, proporcionan herramientas para la creación de materiales interactivos que favorecen el aprendizaje autónomo de los estudiantes, algo muy importante en el contexto del Espacio Europeo de Educación Superior. Todo ello ha permitido desarrollar una gran variedad de recursos docentes en los distintos títulos de la Facultad de Economía y Empresa de la Universidad de Murcia en los que el departamento de Métodos Cuantitativos para la Economía y la Empresa imparte docencia.



El grado en Administración y Dirección de Empresas fue el primero en el que se introdujeron R y RStudio, concretamente en la asignatura de segundo curso dedicada a la estadística inferencial. En el curso 2014-15 se realizaron las prácticas con R y RStudio, de manera experimental, en dos grupos de la asignatura. La experiencia fue tan positiva que, desde el curso siguiente, todos los grupos adoptaron el software. A partir de entonces, y al amparo de distintas convocatorias de innovación docente, se ha confeccionado abundante material de apoyo, aprovechando las posibilidades y recursos html que ofrece R a través de paquetes como rmarkdown. Además, últimamente se está usando de manera intensiva el paquete Shiny para elaborar aplicaciones de carácter interactivo, con las que se pretende reforzar la presentación de contenidos al alumno y evaluar la adquisición de conocimientos y competencias. En este sentido, se está trabajando en la elaboración de ejercicios personalizados para cada estudiante que incorporan feedback en sus distintas etapas de ejecución, lo que permite implementar una evaluación continua de calidad incluso en grupos numerosos. En las asignaturas dedicadas a estadística descriptiva, en primer curso, y econometría, en tercer curso, también se ha preparado material para la realización de prácticas con R y RStudio, que está siendo testado. En el caso de la asignatura de primer curso, en 2017-18 se han realizado, de forma experimental en el grupo bilingüe, un total de cuatro prácticas de gestión de datos, análisis descriptivo y representación gráfica, empleando paquetes como bookdown para la presentación de los materiales. En Econometría las prácticas se resuelven con Eviews, pero se ha desarrollado material para abordar este curso 2018-19 algunas prácticas con R y RStudio en uno de los grupos, de modo que los alumnos puedan valorar las ventajas e inconvenientes de estos dos softwares.

En el grado de Economía se está empleando en 2018-19 el paquete swirl para mejorar la comprensión de conceptos estadísticos y para ejercitar a los alumnos en el uso de distintas herramientas. Las lecciones swirl se desarrollan de forma pausada en la consola de R o RStudio e intercalan texto, imágenes (gráficos generados con R o imágenes de cualquier tipo con fórmulas, esquemas, etc.) y preguntas, guiando al estudiante en su aprendizaje y proporcionando feedback en todo momento. También en este caso se pueden personalizar los ejercicios y utilizarlos en el contexto de una evaluación continua. En el grado en Sociología, de reciente implantación, todas las prácticas de su única asignatura de estadística se han planteado con R y RStudio. Además se han elaborado páginas web explicativas que favorecen el trabajo autónomo del alumno y se han incluido un par de actividades interactivas en la evaluación

continua.

También se usa R en posgrado, concretamente en la asignatura de Métodos Cuantitativos e Indicadores económicos del Máster en Finanzas, de marcado carácter cuantitativo. Se utilizan los paquetes `quantmod`, `fBasics`, `xts` y `rugarch` para analizar series de rendimientos diarios. En particular, el paquete `quantmod` contiene funciones que permiten descargar y manipular datos financieros de las principales fuentes de datos, como son Yahoo Finance y Google Finance, así como diseñar estrategias de negociación. También se introduce a los alumnos al uso del paquete `rugarch` para modelizar el comportamiento de la media y varianza condicionales de las series de rendimientos, lo que es imprescindible para la valoración de derivados y en la gestión del riesgo.

Finalmente, dirigido a alumnos de los últimos cursos que están realizando sus trabajos de fin de grado, así como a alumnos de posgrado o egresados interesados, se está preparando material de trabajo autónomo, básicamente práctico que, mediante ejemplos, ilustra el empleo de técnicas multivariantes. Se trata de un tema muy interesante que, por falta de tiempo, no puede ser abordado en nuestras asignaturas de grado.

A modo de conclusión, nuestra evaluación del proceso de introducción de R en la enseñanza de las asignaturas de estadística del departamento es muy positiva. Los recursos elaborados facilitan el aprendizaje de nociones y técnicas estadísticas usualmente complicadas para los estudiantes de Ciencias Sociales, lo que parece mejorar sus resultados, especialmente el trabajo con actividades interactivas.

Póster 8. 10 años de Jornadas de usuari@s de R en España

Alejandro Cisterna García, Álvaro Hernández Vicente, Ana Belén Marín Valverde, Antonio J Perán Orcajada, Antonio Maurandi López, Aurora González Vidal, Carlos J. Gil Bellosta, Elvira Ferre Jaén, Fernando Pérez Sanz, Francisca Carreño, Francisco López Martínez, Fuensanta Arnaldos Garcia, Isabel Parra Frutos, José Antonio Sánchez Martí, Jose Luis Cañadas Reche, Laura Antón Sánchez, Laura del Río Alonso, Lourdes Molera Peris, Ma Teresa Diaz Delfa, Rafael López Serrano, Ursula Faura Martínez y Virgilio Gomez Rubio.

Un resumen en números de las 10 jornadas de R celebradas desde la primera en Murcia en el año 2009 hasta las décimas (estas celebradas también en Murcia. Se resaltaron las fechas del surgimiento de las diversas asociaciones, la nacional “RHispano” y las agrupaciones locales que vinieron después.

	Murcia 2009	Mieres 2010	Madrid 2011	Barcelona 2012	Zaragoza 2013	Santiago 2014	Salamanca 2015	Albacete 2016	Granada 2017	Murcia 2018
# ponencias	6	9	42		15	25	22	27	26	32
# asistentes	(50)	(20)	132	(70)	97	115	(50)	80	(100)	144
Mesa redonda	Si	Si	no	–	Si	No	No	No	No	No
Posters	No	No	no	–	No	No	No	Si	Si	Si
# talleres	4	2	5	7	4	6	6	6	5	5
Precio de las Jornadas	Gratis	Gratis	Gratis	50-60	10-60	Gratis	Gratis	60-100	Gratis	35-55
Premios	0	0	1	–	3	No	1	1	1	1
Ponencias invitadas	2	1	1	–	6	2	No	2	No	2
Libro de resúmenes	No	No	1	–	Si	Si	No	Si	No	Si
Fotos y videos	No	No	1	–	Si	No	No	No	No	Si
Streaming	Si	No	1	–	no	No	Si	No	No	Si

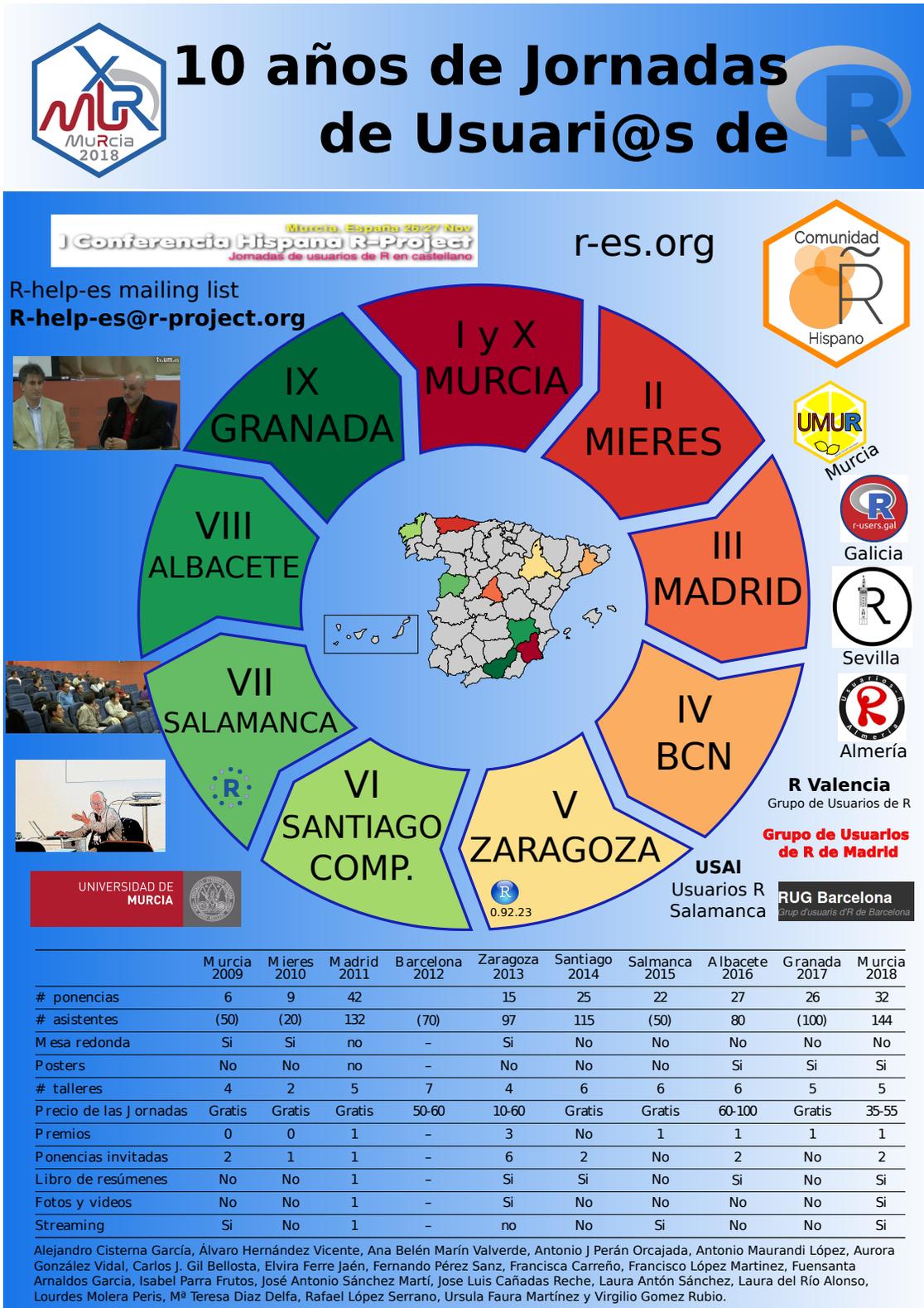


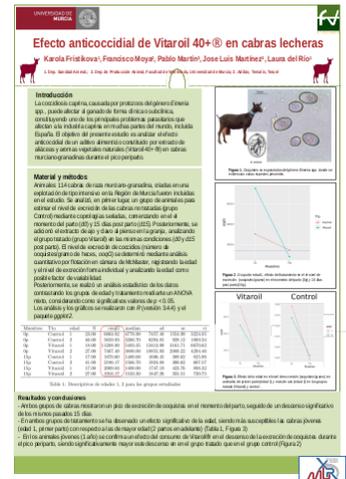
Figura 6.1: El comité organizador de las Xjur.

Póster 9. Efecto anticoccidial del extracto de ajo en cabras lecheras

Karola Fristikova^a, Francisco Moya^b, Pablo Martín^c, Jose Luis Martínez^d y Laura del Río^e

- ^a Dep. Sanidad Animal, Universidad de Murcia
- ^b Dep. de Producción Animal, Facultad de Veterinaria, Universidad de Murcia
- ^c Adibio Tervalis, Teruel
- ^d Adibio Tervalis, Teruel
- ^e Dep. Sanidad Animal, Universidad de Murcia, laurario@um.es

La coccidiosis caprina, causada por protozoos del género *Eimeria*, puede afectar al ganado de forma clínica o subclínica, constituyendo uno de los principales problemas parasitarios que afectan a la industria caprina en muchas partes del mundo, incluida España. En el presente trabajo presentamos datos obtenidos en un ensayo en el que se analiza el efecto anticoccidial de extracto de aliáceas y aromas vegetales naturales (Vitaroil 40(R) en cabras murciano-granadinas durante el pico de periparto. Para ello se han analizado 114 cabras murciano-granadinas, hembras adultas gestantes, procedentes de una explotación intensiva en la Región de Murcia, a las que se les ha suplementando el pienso con el extracto. Sobre estos animales se realizaron coprologías cuantitativas (en cámara de McMaster) cada 15 días, para determinar el nivel de excreción de coccidios en el momento del parto, estimando estos recuentos individualmente y registrando la edad de cada cabra como posible factor de variabilidad. Posteriormente, se realizó un análisis estadístico de los datos contrastando los grupos de edad y tratamiento mediante un ANOVA mixto. En ambos grupos de tratamiento se ha observado un efecto significativo de la edad en la excreción de ooquistes durante el pico periparto, siendo significativamente más alto en las cabras jóvenes (primer parto) con respecto a las de mayor edad (2 partos en adelante). Además, en los animales jóvenes se confirma un efecto protector del Vitaroil (R) disminuyendo significativamente la excreción de ooquistes con respecto al grupo control.



Una versión on-line de buena calidad de los posters puedes encontrarla en <http://gauss.inf.um.es/umur/enlacesxjur.html>

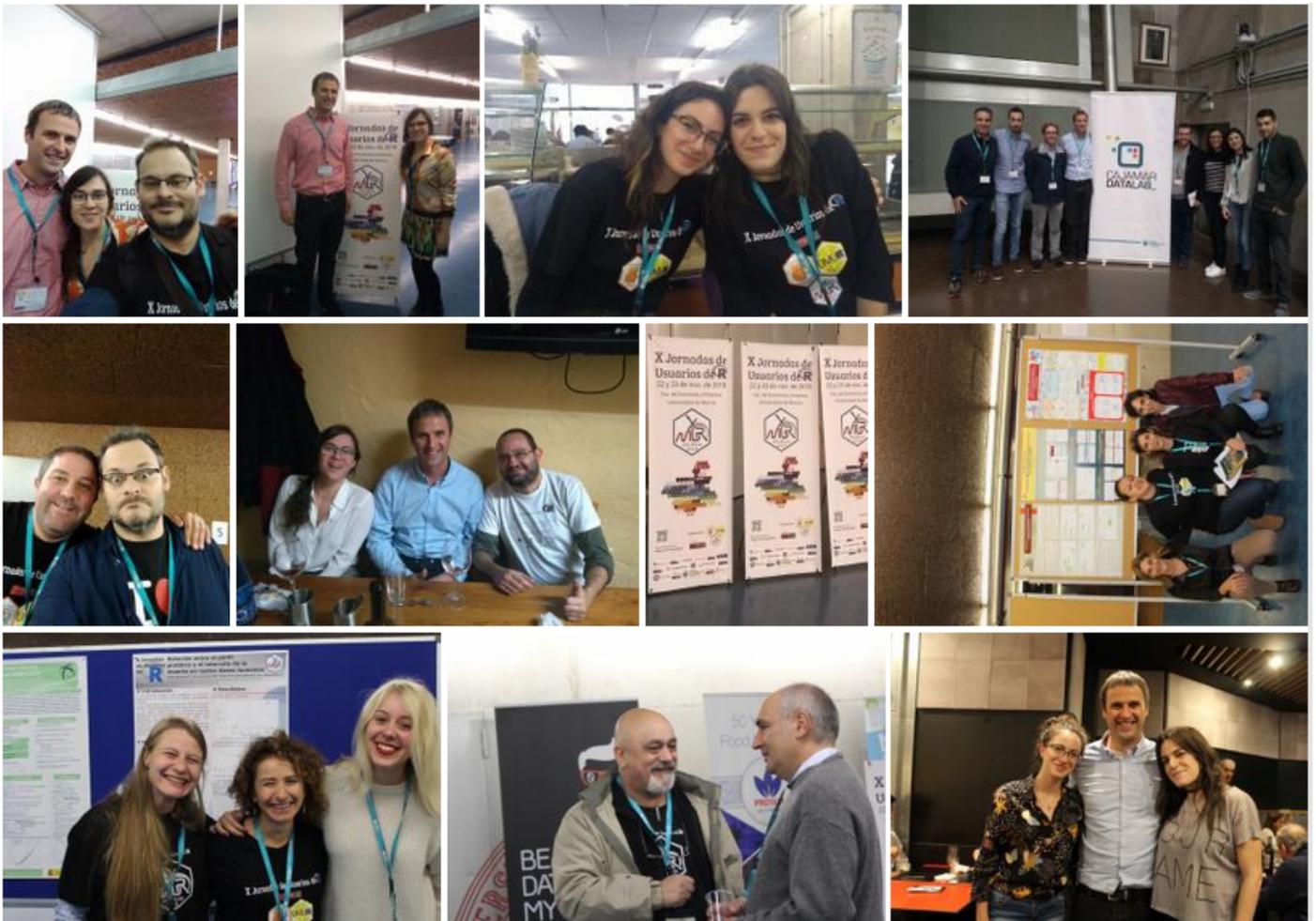
Cuadro 6.1: **Posters**

id	Titulo	Autorías (entidad)
P1	Intelligent Climate. A Machine Learning Solution built on SAP Leonardo & R for HVAC Energy and Comfort optimization	Manuel Jesús Rodríguez Delgado, Antonio Espejo Écija, Christian Fernández Calvillo, Manuel Gutiérrez Plaza, Ismael Moreno Mechbal, Jesús Figueres Cañadas, Jonathan Fuentes Gutiérrez, Alejandro Alija, Mario E. López López
P2	Evaluación del potencial cartográfico de un método de downscaling demográfico de alta resolución	Víctor Martínez García
P3	Relación entre el perfil proteico y el intervalo de la muerte en restos óseos humanos	Gemma Prieto Bonete, María Dolores Pérez Cárceles, Antonio Maurandi López y Aurelio Luna Maldonado.
P4	Aprendizaje usando las lecciones interactivas del paquete swirl de R	Isabel Parra Frutos
P5	Distribuciones tipo fase en memorias RRAM con R	Christian Jose Acal González, Juan Eloy Ruiz Castro y Ana Maria Aguilera Del Pino
P6	Uso de aplicaciones interactivas Shiny para mejorar el rendimiento en las asignaturas de estadística	Juan José Pérez Castejón, Fuensanta Arnaldos García, Ma Teresa Díaz Delfa, Úrsula Faura Martínez y Lourdes Molera Peris
P7	Docencia con R en el departamento de Métodos Cuantitativos para la Economía y la Empresa	Lourdes Molera Peris, Juan José Pérez Castejón, Fuensanta Arnaldos García, Ma Teresa Díaz Delfa, Úrsula Faura Martínez, Isabel Parra Frutos, Susana Álvarez Díez, Francisco Martínez Sánchez y Pepe García Clavel
P8	10 años de Jornadas de usuari@s de R en España	El comité oorganizador local de las xjur
P9	Efecto anticoccidial del extracto de ajo en cabras lecheras	Karola Fristikova, Francisco Moya, Pablo Martín, Jose Luis Martínez y Laura del Río

7

Las Jornadas en imágenes

En http://gauss.inf.um.es/umur/actividades.html#21_x_jornadas_de_usuarios_de_r se puede acceder a una galería de fotos de las jornadas.





8

Repercusión en medios

Antonio Maurandi entrevista en el blog de Máxima Formación
«X JORNADAS DE USUARIOS DE R» CON ANTONIO MAURANDI



¡Ya están aquí las X Jornadas de Usuarios de R del 2018!

Los días 22 y 23 de noviembre en la Universidad de Murcia.

Lourdes Molera, noticia en tv.um.es



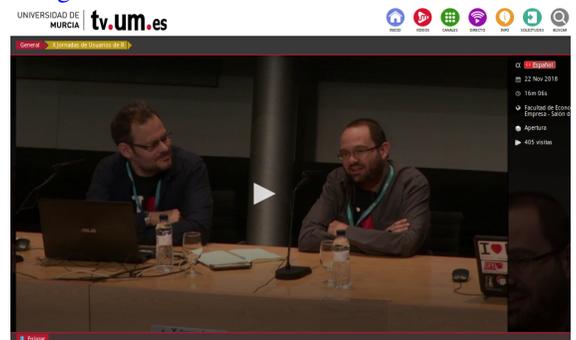
¿Conoces las X Jornadas de Usuarios de R? Hoy han comenzado en la UMU. Han tenido lugar en la facultad de Economía y Empresa

A. Maurandi y F. Javier Ibañez, noticia tv.um.es



Hoy en Canal Umu te presentamos asistir a las 'Jornadas de Usuarios de R'

Inauguración de las X Jornadas de Usuarios de R



Acto de inauguración de las X Jornadas de Usuarios de R

Conferencia plenaria Dr. François Husson

UNIVERSIDAD DE MURCIA | tv.um.es



Conferencia plenaria X Jornadas de Usuarios de R a cargo del Dr. François Husson
"From the package FactoMineR to a project on exploratory multivariate analysis or how to improve the visibility of its package"

Jornadas de Usuarios de R en castellano, 2009 Murcia



8. REPERCUSIÓN EN MEDIOS



*Ya se terminó
Ya se va la gente
Ya sé lo que me vas a decir*

*Que no hay que llorar
Que son cosas que pasan
Y yo siempre lloré por no reír*

*Pero no me queda más memoria
Y no hay foto que quiera borrar*

*Ya se han terminao
las décimas jornadas
ya sé lo que me vas a decir*

*que no hay que llorar
que son cosas que pasan
R te ha hecho muy feliz*

*si has aprendido cosas nuevas,
ve y compártelas con los demás.*

*Ya se termino
Ya es el fin de fiesta
Y nace el tan temido qué dirán*

*Si R se quedará
Si otro software vendrá
Eso solo el tiempo lo dirá*

*si has aprendido cosas nuevas,
ve y compártelas con los demás.*

Letra adaptada de la canción "Fin de Fiesta" de Kevin Johansen.



