
Opiniones sobre la profesión

Statistics in research vs. research in statistics

Inmaculada Arostegui, Irantzu Barrio, María J. Legarreta and
Arantza Urkaregi

Departamento de Matemática Aplicada y Estadística e I.O.
Universidad del País Vasco UPV/EHU

✉ inmaculada.arostegui@ehu.es, irantzu.barrio@ehu.es, mariajose.legarreta@ehu.es,
arantza.urkaregi@ehu.es

Abstract

Celebration of The International Year of Statistics was the main reason to organizing the meeting “Statistics in Research vs. Research in Statistics” at the University of the Basque Country UPV/EHU on 2013 October 9th. The goal was to meet statisticians and statistical users from other scientific areas and to discuss about the field. A main conference about training in statistics and career choices and a round table about statistical consulting were the main activities of the meeting. A poster session and a workshop about sample size completed the program and allowed for active participation of attendees. Participation rate was high, especially among students and researchers. The organizing committee and the participants were satisfied with the result. Discussion of the event was provided by 18 participants, individually or in group.

Keywords: Statistics, Research, Training, Interdisciplinarity

AMS Subject classifications: 62-07, 62Pxx, 92B15

1. Introducción y motivación

Con motivo de la celebración del Año Internacional de la Estadística, el 9 de octubre de 2013 se celebró en la Facultad de Ciencia y Tecnología de la Universidad del País Vasco UPV/EHU una jornada científica que llevaba por título “Estadística en Investigación vs. Investigación en Estadística”. La relevancia de la estadística en la investigación está ampliamente documentada, en especial en las ciencias experimentales y sanitarias (Molenberghs, 2005, Zelen, 2006). Sin embargo, desde principios de la pasada década ya se pone de manifiesto a nivel nacional (Abraira et al., 2001) que la interacción entre estadísticos y profesionales de otras áreas, en la práctica, no es sencilla. Nuestro objetivo al organizar

este evento fue que sirviera como punto de encuentro entre el personal investigador en estadística e investigadores de otras áreas que, bien por interés o por necesidad, utilizaran la estadística en su investigación.

Nos planteamos una jornada abierta fundamentalmente a la participación de investigadores de cualquier área científica; pero, además, nos propusimos atraer otros dos perfiles adicionales. En primer lugar deseábamos promover la asistencia del alumnado, tanto de grado (especialmente de últimos cursos) como de posgrado. Nos parecía importante que el alumnado de matemáticas y estadística se planteara la investigación como una salida laboral, no exclusivamente orientada a la universidad, sino también a centros de investigación y empresas de áreas muy diversas. En este sentido, deseábamos que nuestro alumnado conociera opiniones sobre la profesión, como por ejemplo la recientemente expresada por Lorenzo-Arribas et al. (2013), que les sirvieran como orientación en la elección de su futuro profesional. El segundo colectivo que queríamos atraer era el de las empresas del entorno (en sentido amplio): organismos públicos, centros tecnológicos, etc. . . . En este caso, nuestro objetivo era que el personal que, aun no siendo directamente investigador, en su trabajo estuviera muy cercano o vinculado a la investigación, tuviera conocimiento de los contenidos que la estadística puede aportar a la hora de obtener o mejorar sus resultados. En este sentido, queríamos difundir la idea de que el perfil de la persona experta en estadística puede aportar conocimiento no exclusivamente técnico, sino también científico.

Este manuscrito está dividido en cuatro secciones, de las cuales la primera es esta introducción. La segunda sección aborda la organización y desarrollo de la jornada, incluyendo un breve resumen de las actividades que se desarrollaron en la misma. La tercera sección, que hemos denominado “Otras perspectivas sobre la estadística y la investigación”, recoge las aportaciones de personas participantes en la jornada y ajenas a la organización. Son personas de diferente formación y ámbito laboral que exponen la utilización que hacen de la estadística en sus respectivos ámbitos. Finalmente, en la Sección 4 realizamos una valoración de la jornada, incluyendo nuestro punto de vista como grupo organizador y también las opiniones recibidas de las personas que han participado en la redacción de la Sección 3.

2. Organización y desarrollo de la jornada

El desarrollo de la jornada se planteó en torno a dos actividades principales. La primera de ellas sería una conferencia plenaria por parte de un representante de estadística de una universidad nacional o europea, que se centrara en la relevancia de la formación en estadística orientada a la inserción laboral. La razón de que eligiéramos precisamente el tema de la formación en estadística fue doble. Por un lado, este es un tema que se ha discutido ampliamente en la década anterior (DeMets et al., 2006, Brown y Kass, 2009) y que no estamos muy seguras

de que su importancia sea suficientemente visible en nuestras aulas, tanto desde la perspectiva del profesorado, como desde la del alumnado (Gordaliza, 2013). La segunda razón fue, como hemos comentado anteriormente, que deseábamos atraer la participación del alumnado y vincular la formación con la inserción laboral nos parecía un buen reclamo.

La segunda de las actividades consistiría en una mesa redonda en torno a la asesoría o apoyo en estadística para investigadores de otras áreas científicas. En este caso deseábamos implicar tanto al alumnado, como al personal investigador de otras áreas científicas en un debate sobre necesidades y forma de dar cobertura a las mismas. La mesa redonda se planteó desde dos puntos de vista: uno que representara la oferta de apoyo estadístico a la investigación y otro que se centrara en las necesidades de los/as investigadores/as en este ámbito.

Una vez fijadas las actividades principales de la jornada se decidió incorporar una sesión de posters durante el receso programado a media mañana entre las dos actividades anteriormente mencionadas. El objetivo de la sesión de posters era complementar la jornada con una actividad más participativa por parte de las personas asistentes. De esta forma, se invitaba a toda la comunidad científica interesada en difundir y promover el desarrollo y la aplicación de los métodos matemáticos y estadísticos, así como a todo el personal de la administración pública o empresas interesadas en estas materias, a participar en este evento. Animamos a investigadores del entorno a presentar sus trabajos, tanto en forma de desarrollo teórico, como aplicaciones estadísticas a las ciencias experimentales, sanitarias, sociales, la ingeniería y el arte y humanidades. Asimismo, deseábamos contar con la presencia de temas relacionados con las estadísticas oficiales, la enseñanza de la estadística y el uso y desarrollo de software estadístico.

Por último, por sugerencia de una de las investigadoras participantes en la mesa redonda, decidimos completar la jornada con una actividad formativa por la tarde. Tras consultar con algunos/as investigadores/as de nuestro entorno más cercano sobre posibles temas de interés, especialmente estudiantes de posgrado, obtuvimos una clara preferencia por temas relacionados con la selección de la muestra y el cálculo del tamaño muestral. De esta forma, el programa de la jornada se completó con la oferta de un taller titulado “Tamaño muestral: ¿cómo estimar n adecuadamente?” de acceso limitado únicamente a 30 de los asistentes a la jornada, previa inscripción y proceso de selección.

La Tabla 1 muestra información detallada sobre las personas inscritas en la jornada, clasificada según formación académica, procedencia y vinculación laboral. Los posters presentados por área temática, en base a la clasificación realizada por los/as autores al enviar el resumen, se recogen en la Tabla 2.

	Nº de personas inscritas
Total	195
Formación académica	
Matemáticas	91
Ciencias Experimentales	33
Ciencias de la Salud	19
Ciencias Sociales	13
Sin información	39
Vinculación laboral	
Estudiantes	83
PDI	43
Investigadores	50
Otros	19
Procedencia	
UPV/EHU	115
Otros	80

Tabla 1: Personas inscritas en la jornada por formación académica, procedencia y vinculación laboral.

	Nº de pósters
Área temática	
Aplicaciones estadísticas en Ciencias de la Salud	13
Aplicaciones estadísticas en Ciencias Experimentales	5
Aplicaciones estadísticas en Ciencias Sociales	1
Aplicaciones estadísticas en Ingeniería	1
Enseñanza de la Estadística	2
Estadística oficial	2
Investigación Operativa	2
Metodología estadística	4
Software estadístico	2
Teoría de probabilidades	1
Otras	3
Afiliación	
UPV/EHU	13
Otras	23

Tabla 2: Posters presentados por área temática.

2.1. La formación de doctores en Estadística orientada a la inserción laboral: algunos ejemplos

La conferencia plenaria corrió a cargo del profesor Pere Puig, catedrático de Estadística de la Universitat Autònoma de Barcelona. El profesor Puig comenzó la conferencia hablando del doctorado en general, puntualizando que la inserción laboral de doctores era una asignatura pendiente en este país, que había una necesidad de ofrecer una formación multidisciplinar a los/as doctorandos/as y que había que implicar a las empresas e instituciones en la formación de doctores. Como estrategia de inserción laboral, enumeraba las ventajas que un doctor/a podía ofrecer a las empresas. Como posible propuesta, planteaba el doctorado “industrial”, un proyecto de investigación estratégico de una empresa, donde el doctorando/a desarrollara su formación investigadora, en colaboración con una universidad, y que ésta fuera objeto de una tesis doctoral.

A continuación, pasó a hablar del doctorado en estadística en particular, el cual, comparado con otras áreas científicas, es relativamente fácil de integrar en proyectos de investigación multidisciplinarios. Muchos grupos de investigación en ciencias experimentales, sociales y salud necesitan estadísticos en sus equipos, y si la persona experta en estadística es doctor, esto podría suponer una ventaja favorable para conseguir proyectos competitivos. En su opinión, habría que conseguir que el/la estadístico/a fuera un miembro más del grupo de investigación, no sólo un técnico de soporte a la investigación.

Finalmente, presentó varios ejemplos de proyectos de tesis doctoral en estadística consolidados en proyectos de investigación multidisciplinarios, dirigidos o codirigidos de manera conjunta por miembros de varias instituciones e integrados en alguna empresa o institución fuera de la universidad.

2.2. Apoyo estadístico a la investigación. ¿Qué demandan los/as investigadores/as?

El objetivo de la mesa redonda fue reflexionar sobre la utilización de los métodos estadísticos en diferentes campos de investigación desde una doble perspectiva, quienes ofertan apoyo estadístico y quienes son usuarios/as de la estadística en su investigación.

En primer lugar, se abordaron diferentes formas de dar apoyo estadístico a la investigación. Maribel Arriortua, Directora de los Servicios Generales de Investigación (SGIker) de la UPV/EHU explicó el funcionamiento de estos servicios, en los que no existe una unidad específica de estadística, aunque sí se han impartido cursos y se ha colaborado en algunos proyectos aplicando métodos de quimiometría. Anna Espinal, Coordinadora del Área de Investigación del Servei d'Estadística Aplicada de la Universitat Autònoma de Barcelona, expuso la amplia experiencia de este servicio, que ofrece, desde el año 2000, asesoría estadística a toda la comunidad universitaria e investigadora, y también a empresas de dife-

rentes ámbitos de conocimiento. Por último, Vicente Lustres, director gerente de Biostatech, S.L., spin-off creada en marzo de 2012 a partir del Grupo Interdisciplinar de Estadística, Computación, Medicina y Biología (GRID[ECMB]) de la Universidad de Santiago de Compostela, explicó las fases de creación de una spin-off y las ventajas que plantea de cara a la conexión del medio investigador y el empresarial o a la creación de empleo.

En la segunda parte de la mesa redonda participaron tres investigadores de las áreas de Química, Biología y Medicina. Nestor Etxebarria, Director del Departamento de Química Analítica e investigador del Grupo IBeA (Ikerkuntza eta Berrikuntza Analitikoa - Investigación e Innovación Analítica) destacó que muchas de las operaciones y tareas que se realizan en Química Analítica necesitan de metodología estadística, por lo que recalcó la necesidad de una formación sólida en estadística. También puso de manifiesto las dificultades en la colaboración entre estadísticos y químicos (analíticos), señalando que muchos químicos analíticos realizan su investigación en el desarrollo de herramientas quimiométricas. Desde su punto de vista, “en esa posible colaboración está casi todo por hacer, y seguramente todos tenemos mucho que ganar”.

Miren P. Cajaraville, investigadora principal del Grupo “Cell Biology and Environmental Toxicology” afirmó que, en su área, “no se puede publicar ningún resultado que no haya sido sometido a análisis estadístico”. Destacó los problemas con los que se encuentran a la hora de la aplicación de herramientas estadísticas debido a la gran variabilidad en los procesos, sistemas y respuestas biológicas. Finalmente, abordó la diferencia entre significación estadística y significación biológica. En ocasiones, nos encontramos con resultados que son estadísticamente significativos aunque no lo son biológicamente. Al contrario, en otras ocasiones no se obtiene significación estadística para una tendencia con relevancia biológica, lo cual puede deberse a que el tamaño muestral es pequeño.

José M. Quintana, responsable de la Unidad de Investigación Clínica del Hospital Galdakao-Usansolo, desgranó las demandas de los investigadores clínicos al personal bioestadístico, tanto en el diseño del estudio como en la selección y ejecución del método de análisis óptimo y en la interpretación y explotación de los resultados. Para este investigador, una labor fundamental del personal bioestadístico es saber interpretar y explicar lo que los investigadores realmente están buscando, es decir, ser capaces de traducir el lenguaje sanitario a un lenguaje estadístico. Incidió en el aprendizaje mutuo entre investigadores clínicos y bioestadísticos al desarrollar un trabajo multidisciplinar. El profesor Quintana insistió en la necesidad de una formación más específica y especializada de los/as estadísticos/as, para así facilitar su integración en diferentes áreas de investigación.

2.3. Tamaño muestral: ¿Cómo estimar n adecuadamente?

El taller fue impartido por Vicente Lustres y Mónica Ratón de Biostatech. Consistió en una introducción sobre la importancia del cálculo del tamaño de una muestra a la hora de abordar un experimento, para aportar mayor validez y calidad científica a nuestros resultados y conclusiones. La realización de un cálculo a posteriori puede dar como resultado una muestra inadecuada, y puede implicar un incremento en los costes de ejecución y dedicación para dar solución al problema planteado. Para ello, los profesionales e investigadores del ámbito de las ciencias de la vida, no siempre disponen de las herramientas necesarias para este cálculo, o la complejidad del estudio supone una importante limitación en el desarrollo de esta tarea.

Se impartieron nociones básicas sobre el cálculo del tamaño muestral, en situaciones como la evaluación de una proporción o una media, o como la comparación de dos medias o de dos proporciones, tanto en muestras independientes como dependientes. Mediante diferentes ejemplos se vieron varios casos de problemas de estimación de tamaños muestrales y cómo abordarlos, mostrando la importancia de los parámetros a fijar y las condiciones que se deben verificar, para obtener una muestra representativa y poder extrapolar los resultados a la población objeto del estudio, con menores márgenes de error.

Adicionalmente, se mostraron diferentes alternativas de uso de software con los que poder calcular el tamaño muestral. En concreto, se realizaron varios ejemplos utilizando el software G*Power.

3. Otras perspectivas sobre la estadística y la investigación

En esta sección incorporamos las aportaciones realizadas por algunos/as de los asistentes en cuanto al uso de la estadística como herramienta fundamental en su campo de investigación. En total han sido siete las aportaciones recibidas, algunas individuales y otras grupales, las cuales incluimos con el nombre de sus autores.

3.1. Eustat–Euskal Estatistika Erakundea y la investigación estadística

Jorge Aramendi Rique, Elena Goñi Rementería, Ines Garmendia Navarro, Imanol Montoya Arroniz y Anjeles Iztueta Azkue

Eustat-Euskal Estatistika Erakundea

Eustat, Organismo Autónomo del Gobierno Vasco, acudió a la jornada científica, principalmente con la intención de compartir dos de sus proyectos de investigación y desarrollo en este foro.

El interés de Eustat en la investigación no es nuevo. Desde hace tiempo, Eustat promueve iniciativas investigadoras de diferentes tipos, en particular con

las Becas de Formación e Investigación Estadístico–Matemática, los Seminarios Internacionales de Estadística y los distintos proyectos de investigación en colaboración con departamentos de la universidad.

Los dos proyectos de I+D presentados, y que actualmente se están desarrollando en nuestro programa de becas, son

- Depuración Selectiva de Bases de Datos, cuyo objetivo es definir y aplicar una metodología para optimizar la fase de depuración de la información estadística, que puede ser muy costosa en tiempo y en recursos.
- Enlace de Encuestas o *Statistical Matching*, con el objetivo de aplicar la técnica que permite enlazar estadísticamente, a través de características o modalidades comunes, información proveniente de encuestas independientes.

Ambos temas reflejan intereses comunes de la Estadística Oficial Europea, con aplicaciones en muchas operaciones estadísticas, ya sean basados en encuestas o en registros administrativos.

La jornada resultó un foro adecuado para el intercambio de experiencias en el ámbito de la estadística y la investigación, a través de la conferencia y la mesa redonda, y de la sesión de pósters. En la primera parte se trataron problemas y carencias de formación, así como la importancia de la estadística en el proceso científico y en el desarrollo de la investigación. Desde diferentes ámbitos se subrayó la necesidad de personal con capacitación en esta disciplina, que además debería ser materia común para todos los programas de posgrado y especialización.

En la segunda parte se brindó la oportunidad a organismos y equipos de investigación de compartir proyectos que responden a necesidades reales y que están fundamentados en metodología estadística. La experiencia resultó interesante en particular por la diversidad de aplicaciones que se presentaron.

En conclusión, valoramos el evento de forma muy positiva ya que Eustat pudo estar en contacto con el mundo universitario e investigador, recogiendo ideas y aportando nuestro granito de arena. El intercambio de conocimiento y experiencias forma parte de nuestro interés por la innovación y la calidad en la producción estadística.

3.2. Las matemáticas de AZTI

Agurtzane Urtizberea, Dorleta García, Leire Ibaibarriaga y Sonia Sanchez
Azti-Tecnalia, Unidad de Investigación Marina, Sukarrieta

AZTI-Tecnalia es un centro tecnológico experto en investigación marina y alimentaria en el País Vasco, y como tal, es lugar de trabajo de numerosos licenciados/as en ciencias. La Unidad de Investigación Marina incluye a 4 licenciadas

en matemáticas por la UPV/EHU. Su trabajo se centra principalmente en estadística aplicada, siempre dentro de grupos multidisciplinares, interaccionando con biólogos, físicos y economistas entre otros. La interacción con los diferentes departamentos de la universidad se considera prioritaria en un centro tecnológico, de forma que AZTI-Tecnalia intenta favorecer de forma activa todas aquellas iniciativas que potencien esa relación.

La estadística es una ciencia de carácter muy aplicado y en AZTI-Tecnalia se utiliza sobre todo como una herramienta para investigar en otros campos. Pero es necesario recalcar que sin avances en la investigación en estadística teórica no se puede avanzar en nuevas aplicaciones y metodologías.

Jornadas como ésta son una buena ocasión para mostrar algunos de los ejemplos de cómo y para qué usamos la estadística en investigación marina a un público que no suele ser el habitual en nuestro ámbito (con una mayor formación metodológica). Del mismo modo nos permite enseñar al alumnado lo que hacemos para que así puedan considerar la opción de trabajar, estudiar un máster o realizar una tesis en Azti-Tecnalia. La participación en la sesión de posters fue muy alta y por tanto, una buena oportunidad de conocer los trabajos que se están llevando a cabo tanto en la universidad, como centros de investigación o empresas, permitiéndonos encontrar contactos para futuras colaboraciones.

En general, este tipo de jornadas y talleres puede ser un método clave para crear una interacción más frecuente, a veces inexistente, entre las universidades, centros tecnológicos y empresas. Asimismo, no deberían de tener lugar solo dentro de celebraciones excepcionales como este año, sino con un carácter más periódico y de un modo más fluido entre los diferentes agentes implicados.

3.3. Muestreos de campo y estadística: temas de difícil comunión

Aitor Larrañaga

Departamento de Biología Vegetal y Ecología, Facultad de Ciencia y Tecnología, UPV/EHU

Egoitz Salsamendi

Departamento Zoología y Biología Celular Animal, Facultad de Ciencia y Tecnología, UPV/EHU

Como biólogos la única formación que recibimos sobre estadística fue allá en el primer curso de la licenciatura. Quienes atendían en clase estudiaron, entre otras cosas, la teoría de la probabilidad, la normalidad, la tendencia central de los datos y los requerimientos para realizar un tipo u otro de análisis estadístico. Convenientemente, los ejemplos que se trabajaban no dejaban ni un ápice de duda sobre la técnica adecuada para cada situación: ¡la estadística era una herramienta sencilla de entender y simple de utilizar!

Tras licenciarnos ambos decidimos emprender nuestra andadura en el estimulante mundo de la investigación en la Universidad del País Vasco UPV/EHU.

Aunque de departamentos distintos, nuestras líneas de investigación conllevan mayoritariamente muestreos de campo, con lo que nos topamos con inconvenientes similares a la hora de aplicar la estadística en nuestras investigaciones. Ambos coincidimos en que existe un abismo entre lo que debemos (lo que nos dicta la estadística para obtener un resultado robusto) y lo que podemos (lo que nos dicta el trabajo de campo) hacer.

Nos explicamos con casos. Supongamos que queremos conocer cómo afecta la presencia de la trucha al funcionamiento de los ríos. Para responder a esta pregunta podemos comparar un número de variables entre ríos con presencia de trucha y ríos sin trucha y, además, que estos ríos no difieran en el resto de los factores. En este caso, el estudio de cada sujeto (río) puede ser tan tedioso que no podamos abordar más ríos (por falta de fondos, medios, tiempo, etc.). Con suerte, al final terminaremos con 4-5 ríos sin trucha y otros 4-5 con trucha; tamaño muestral que pondría nervioso a cualquier estadístico. Por otro lado, supongamos que queremos saber cuál es el efecto de la presencia de montañeros en el tamaño del área de campeo de los osos en el Pirineo. Con una población estimada de 18-27 individuos, el tamaño muestral es excesivamente reducido para unos resultados estadísticamente satisfactorios. Es este tipo de problemática con la que se enfrenta cualquier biólogo de campo en sus investigaciones diarias.

En estos casos tenemos que recurrir a técnicas estadísticas que son menos ortodoxas, y muchas veces tenemos que adaptar la estadística a los datos (y no al revés como sería recomendable) para poder extraer información objetiva de ese escaso pero importante pool de datos. Es este tipo de problemática la que nos ha llevado a sumergirnos en la estadística y controlar aspectos más especializados. Pero además, hemos “forzado” a estadísticos profesionales a pensar en los datos disponibles obtenidos en el campo y plantear alternativas correctas (¡o menos malas!) para abordar la investigación en cuestión. Esta interacción entre estadísticos y biólogos de campo resulta indispensable para llevar a cabo (cada vez más) investigaciones científicas y es beneficiosa para ambas partes. Nosotros obtenemos una base importante para poder responder con objetividad a las preguntas que nos habíamos planteado, pero por otro lado, las personas expertas en estadística tienen la posibilidad de explorar nuevos campos de estudio en base a la tipología de datos que obtenemos. Es de esta fusión interdisciplinar de donde se genera el conocimiento que marca el desarrollo de las líneas científicas. A este respecto, iniciativas como esta jornada científica son herramientas para dar a conocer esta fusión interdisciplinar, y a la vez potenciarla.

3.4. Observatorio Turístico de Euskadi: nuevas formas de medir el turismo y la movilidad

Mikel Zurutuza

CICTOURGUNE Cooperative Research Center in Tourism

La jornada fue una experiencia enriquecedora desde el punto de vista tanto de los asistentes, investigadores (posgrados y doctorandos) y estudiantes de universidad, así como también para docentes y representantes del mundo laboral presentes en la mesa.

Nuestro principal objetivo en esta jornada era dar a conocer los recientes trabajos de investigación que estamos realizando en CICtourGUNE en relación con el sector del turismo y el análisis de datos, y por otro lado, también establecer contacto con personas que trabajan en los diversos ámbitos científicos y poder conocer los métodos estadísticos y técnicas de análisis que se usan en las distintas problemáticas de cada una de ellas.

En cuanto al desarrollo de la jornada, destacar principalmente el intercambio de opiniones y puntos de vista que surgió en la mesa redonda organizada a la mañana, positiva en cualquier sentido para conocer las inquietudes y requerimientos de los diferentes sectores laborales en cuanto a la necesidad del uso de la estadística y el análisis de datos. Cabe destacar principalmente la cuestión más importante que surgió en el debate: el choque interdisciplinar entre las personas involucradas en la materia científica a estudiar y el estadístico.

Como reflexión final, comentar la relación existente a día de hoy entre la estadística y la investigación en turismo y movilidad de Euskadi. El turismo es, por un lado, un sector transversal, y por esta razón se requiere la extracción y utilización de datos de diversa naturaleza sectorial. Por otro lado, siendo el turismo un sector intensivo en información, existen varias fuentes de datos oficiales (institutos estadísticos) y secundarias (operaciones estadísticas a medida) que tratan de describir la coyuntura turística. Además, existe una dificultad añadida a la hora de trabajar con estadística a nivel sub-nacional o regional. Las operaciones que se realizan a nivel nacional no se ajustan en su totalidad a la realidad de las distintas regiones, u ofrecen unos datos genéricos que necesitan complementarse con operaciones secundarias a medida.

Una de las líneas estratégicas del centro se enfoca en la medición y el análisis de los datos del turismo. En este sentido nace un sistema de información que unifica todos estos datos de distinta naturaleza y se integran en una plataforma que ofrece capacidades de análisis estadístico y visualización: el Observatorio Turístico de Euskadi. Se han obtenido datos estadísticos tradicionales de demanda y oferta turística tales como alojamiento turístico, entradas en museos, petición de información en oficinas de turismo, ... Como complemento se recaban datos de transporte tales como vuelos y cruceros, y se combinan con datos de distintas operaciones estadísticas regionales como el comportamiento de los turistas

en Euskadi y otras a nivel estatal como los movimientos de los españoles en fronteras y el gasto de los turistas extranjeros.

Por otro lado, y como complemento a la estadística más tradicional, se está tratando de abarcar otro tipo de fenómenos relacionados con el destino con proyectos como eGIStour (se pretende medir y modelar los patrones de movimiento en los diferentes puntos de interés, POI - *point of interest*), Dynamic Pricing (se analiza la variabilidad de la oferta hotelera), Monitor Web de Destinos (se observa y modeliza el comportamiento online en las webs de promoción turística) o Monitor de *Social Media* (se tienen en cuenta todas las experiencias, opiniones, comentarios, valoraciones, ... que la gente aporta en las redes sociales de Internet).

La medición y el análisis de todas estas materias resulta estratégico para poder dar una visión holística de la coyuntura del sector y en definitiva mejorar la competitividad de un sector como el turismo en Euskadi. Para ello la estadística juega un papel fundamental.

3.5. La estadística en las actividades del grupo de investigación RSAIT

Itziar Irigoien y Yosu Yurramendi

Departamento de Ciencias de la Computación e Inteligencia Artificial, Facultad de Informática UPV/EHU

Queremos empezar reconociendo que esta jornada contribuyó notablemente a visualizar la presencia de la estadística en muy diferentes ámbitos y por ello queremos felicitar a las organizadoras y agradecer la invitación que nos han ofrecido para colaborar en este trabajo. Aprovechamos la ocasión para comentar los aspectos en los que utilizamos la estadística en nuestro grupo de investigación RSAIT (*Robotika eta Sistema Autonomoen Ikerketa taldea*) en la Universidad del País Vasco UPV/EHU. Nuestro grupo trabaja principalmente en el área de robots móviles, por ello, el trabajo en aprendizaje automático y métodos estadísticos es importante. Por ejemplo, está todo el campo conocido como *Probabilistic Robotics*, que trata de modelar la percepción del robot a través de los sensores teniendo en cuenta la incertidumbre. Sin embargo, no es el único ámbito en el que se utiliza la estadística ya que con el objetivo de dotar de diferentes capacidades al robot son muchos los ámbitos en los que hay que trabajar con “datos”.

Como caso particular, comentamos el problema de la exploración autónoma del robot y nuestra experiencia de cómo lo abordamos. Por exploración autónoma entendemos que el robot pueda moverse en un entorno desconocido “sin perderse” y pueda proporcionar al final del proceso un mapa topológico del entorno explorado. En este caso, el entorno era el edificio de la Facultad de Informática de la UPV/EHU, precisamente la tercera planta del mismo. La estructura de la planta donde se movía el robot es del tipo pasillos-oficinas, es decir, en la

planta hay varios halls y pasillos que dan acceso a los despachos y laboratorios de la facultad. Por tanto, el robot se encontraba en entornos tipo “pasillo”, “cruce de pasillos” o “halls”. Con la información del sistema odométrico se dotó el robot con la capacidad de distinguir entre los tres tipos de entornos, pero una vez identificaba el robot que se encontraba, por ejemplo, en un pasillo, no sabía distinguir si ya había pasado por ese mismo pasillo o se trataba de uno nuevo.

En la exploración autónoma, este problema se conoce como el *loop-closing*. Sin embargo, desde el punto de vista estadístico, se puede abordar como un problema de “tipicalidad”. Es decir, dados C_1, \dots, C_k clases y una nueva observación y_0 , antes de clasificar y_0 en alguno de estos C_1, \dots, C_k , decidir si y_0 proviene de alguna de estas clases o bien de una clase desconocida. De esta manera, el robot puede primero decidir si el pasillo en el que se encuentra es nuevo o no, y en caso de que no sea nuevo pueda decidir en cuál de los pasillos que previamente ha visitado se encuentra. En este caso particular, estudiar y trabajar en el desarrollo de nuevos métodos estadísticos nos dio la oportunidad de dar una solución al problema del *loop-closing*.

Sin embargo, no es el único problema en el que el análisis de datos, el aprendizaje automático y la estadística son una herramienta fundamental. El estudio en tratamiento de imágenes es también un ámbito importante ya que la cámara puede ser el ojo del robot. Además, hay que integrar la información proveniente de diferentes tipos de fuente como pueden ser láseres, kinect, etc. Todo ello es fundamental para poder buscar métodos con los que el robot pueda decidir si en su camino hay un obstáculo inmóvil, una presencia humana o cierto objeto determinado a su alcance. Incluso, para poder avanzar en la interacción persona-computador, resulta necesario tener que trabajar en el tratamiento de voz y en el procesamiento del lenguaje natural. Para seguir avanzando en todo ello, es necesario desarrollar métodos estadísticos adecuados para cada ámbito. Es nuestro propósito seguir trabajando en ello.

3.6. La estadística en el mundo biosanitario

Urko Agirre, Ane Anton y Anette Unzurrunzaga

Unidad de Investigación Clínica, Hospital Galdakao-Usansolo

La labor de un Bioestadístico se centra principalmente en resolver problemas del ámbito biosanitario a través del análisis de datos, para verificar o formular nuevas hipótesis. Por eso, y más allá de la búsqueda de la significación y el cálculo del tamaño muestral, las personas expertas en la materia debemos de ejercer de enlaces entre el lenguaje médico y el matemático, ya que es necesario para poder interpretar correctamente y de una manera crítica los resultados obtenidos. Un buen trabajo de investigación en salud debe dedicar un espacio a describir los métodos de análisis de datos utilizados, indicando los factores determinantes para un buen análisis en cuanto al tipo de diseño y el conocimiento de las variables

de estudio.

Una de las problemáticas más frecuente en este ámbito es la malinterpretación entre la significación clínica y estadística. La mayoría de los resultados que provienen de proyectos de investigación se basan en la significación estadística, con el fin de evaluar las diferencias o asociaciones que se dan entre grupos o variables de interés. Planteando el contraste de hipótesis adecuado y llevando a cabo los análisis estadísticos pertinentes, obtenemos un p -valor. Dicho p -valor, bajo un umbral significativo establecido (α o error Tipo I), indica la presencia o no de una significación estadística.

De acuerdo a nuestra experiencia personal en el área biosanitaria, a menudo los investigadores clínicos centran su atención en saber la significación estadística de sus objetivos principales ($p < \alpha$). Sin embargo, el p -valor, tal y como hemos apuntado previamente, da información limitada. Es decir, solo indica si es significativo o no, pero no demuestra cómo de importante es el resultado de este análisis estadístico.

Además, en la investigación clínica no solo es crucial evaluar las diferencias estadísticamente significativas entre diferentes grupos, sino también es relevante, siempre que sea posible, medir cómo de importante es el resultado (por ejemplo, medir la efectividad y eficacia de una intervención quirúrgica). Y la significación estadística no nos proporciona información acerca de la relevancia clínica. Es por ello que en el ámbito clínico se tiende a malinterpretar estos dos conceptos. Por un lado, estudios con tamaño muestral grande pueden derivar en resultados estadísticamente significativos, pero tamaño de efecto pequeño (no relevancia clínica). Por otro lado, en estudios con tamaño muestral pequeño, la malinterpretación se refleja cuando obtenemos un resultado estadísticamente no significativo con un tamaño de efecto grande. Por ello, se han desarrollado varios métodos para medir la relevancia clínica, como son, la diferencia mínima importante, tamaño del efecto, etc. Ambas significaciones se pueden considerar complementarias a la hora de presentar resultados clínicos. Los investigadores deberían ir más allá de la significación del p -valor.

Este es uno de los claros ejemplos donde se manifiesta la necesidad de la colaboración entre profesionales de distinto ámbito (bioestadísticos-clínicos). En esta jornada, pudimos comprobar que el trabajo multidisciplinar compuesto por especialistas en el área y estadísticos es fundamental para el buen desarrollo de un trabajo de investigación.

3.7. Técnicas estadísticas no estándar para el análisis de series temporales ambientales

Josue Polanco

BC3 - Basque Center for Climate Change

En la actualidad el uso de la estadística en la investigación científica es una

herramienta prácticamente imprescindible en todas las áreas de la ciencia y en lo que respecta a las ciencias ambientales no es una excepción. Las técnicas estadísticas han sido tradicionalmente utilizadas para el análisis de series temporales ambientales (climáticas, meteorológicas, oceanográficas, etc.). Sin embargo, debido a la naturaleza propia de estas series temporales (e.g., contienen mucho ruido, suelen ser cortas, no equiespaciadas temporalmente, presentan auto-correlación, persistencia, estacionalidad, tendencia, eventos multi-periodicos, etc.), así como a las características de los sistemas ambientales (no están aislados, presentan no-linealidad, los procesos físicos, químicos, biológicos y geológicos que tienen lugar en estos sistemas pueden ocurrir en diferentes escalas espacio-temporales, etc.), los especialistas que analizan estos tipos de datos han tenido que adaptar, implementar, incluso desarrollar ciertas técnicas estadísticas “no estándar” para poder analizarlos adecuadamente. Esto, con el objeto de obtener información cuantitativa que nos permita ganar conocimiento de los fenómenos ambientales a estudio (Mudelsee, 2010, Von Storch & Zwiers, 1999).

Entre las técnicas estadísticas no estándar más utilizadas para el análisis de series temporales ambientales destacan las técnicas de análisis espectral mediante la transformada de Fourier. Estas técnicas están diseñadas para analizar datos de tipo estacionario, pero las series ambientales no siempre satisfacen este requerimiento. Una herramienta que enfrenta el inconveniente de la no estacionariedad, es el análisis espectral mediante la transformada continua de wavelet. Actualmente, existen varias implementaciones metodológicas, con sus respectivos desarrollos computacionales, diseñadas específicamente para realizar análisis espectral de wavelet vía la transformada continua a series temporales ambientales (una breve revisión de esta técnica y algunas aplicaciones en ciencias ambientales fue presentado por Polanco (2011)).

Por otra parte, la forma más común de enfrentar el hecho de que una serie no es equiespaciada, es interpolar la serie original y convertirla en una serie equiespaciada. Sin embargo, la interpolación, independientemente de la técnica utilizada, puede afectar la estimación del espectro (sobrestimando en las bajas frecuencias y subestimando en las altas). Por tanto, en ciencias ambientales, es relativamente común utilizar una técnica desarrollada e implementada computacionalmente (lleva de nombre REDFIT, es de uso libre y está disponible en internet), que trabaja directamente con las series temporales ambientales no equiespaciadas evitando aplicar cualquier técnica de interpolación, y al mismo tiempo tiene en cuenta el ruido “rojo” en la estimación del espectro (Schulz & Mudelsee, 2002).

Para poder llevar a cabo con éxito estas tareas, es habitual contar en los equipos de investigación con personal con amplia experiencia o con una sólida formación en estadística, recurrir a servicios profesionales de asesoría estadística o participar constantemente en eventos académicos relacionados con la estadística y sus aplicaciones en la investigación científica como esta jornada. Desde mi

punto de vista, este tipo de eventos podrían también ser la semilla de "escuelas o talleres de trabajo" para abordar problemáticas comunes o específicas en la investigación científica donde la estadística juega un papel fundamental, lo cual podría ser beneficioso no sólo para el personal investigador, sino también para la propia UPV/EHU.

4. Valoración y conclusiones

La valoración que desde la organización hacemos de la jornada es muy positiva. Queremos destacar, en primer lugar, la aceptación y gran participación en la jornada, no solo por la comunidad universitaria, sino también por parte del entorno investigador y tecnológico, superando nuestras previsiones iniciales. Además de la cantidad de asistentes, queremos también destacar la variedad de ramas de conocimiento. Si bien la mayoría de las personas asistentes procedían de las ciencias experimentales y las sanitarias, las ciencias sociales y jurídicas (especialmente sociología y psicología) también estuvieron presentes. Investigadores de diferentes ámbitos dieron su punto de vista y compartieron sus experiencias laborales en cuanto a la necesidad de utilizar herramientas estadísticas para poder llevar a cabo sus respectivos trabajos. En las valoraciones recibidas desde diferentes participantes hay una coincidencia general en un tema que surgió en varias ocasiones a lo largo de la jornada: la necesidad de formación en estadística del personal investigador de una gran variedad de áreas científicas.

Básicamente fueron dos las ideas que centraron los debates de esta jornada, la formación y la colaboración. La formación especializada en estadística, además de ser el tema de la conferencia plenaria, fue una de las ideas que más se repitió durante la jornada, tanto por los componentes de la mesa redonda, como por varios participantes del público. Por un lado, se planteó la necesidad de una buena formación estadística en las titulaciones de otras áreas de conocimiento. Mientras que, por otro lado, se puso también de manifiesto que la formación de los/as matemáticos/as recién llegados al mundo laboral en el área de la estadística no era suficiente, y se hizo hincapié en que una formación teórica, junto con una formación práctica, sería una buena alternativa ante este problema. El otro punto que centró el debate fue, sin duda, la importancia de la colaboración entre investigadores/as del área de estadística e investigadores/as de otras áreas de conocimiento, potenciando el trabajo multidisciplinar. Una colaboración que, en muchas ocasiones, no estaba exenta de dificultades, pero que sin duda debía ser enriquecedora para ambas partes.

Las diferentes presentaciones que tuvieron lugar en la jornada ayudaron en gran medida al alumnado presente, a mostrar las diferentes líneas de trabajo en las que, en un futuro próximo, puedan aplicar sus conocimientos. La sesión de posters permitió visualizar con resultados reales la aplicabilidad de la estadística en diferentes ámbitos (educación, biología, ámbito sanitario, química, economía,.

. .) y, a su vez, los investigadores participantes podían conocer nuevas técnicas para luego poder aplicarlas en sus propios trabajos de investigación, difundiendo y promoviendo así, el desarrollo y la aplicación de los diferentes métodos matemáticos y estadísticos. Sin embargo, muchos participantes comentaron que el tiempo que se dió para ver y presentar los posters, fue insuficiente para poder intercambiar conocimientos con los demás expositores.

El taller “Tamaño muestral: ¿Cómo estimar n adecuadamente?” fue valorado como muy interesante y acertado, debido a su carácter práctico y su potencial de aplicación a numerosos campos. Tanto la teoría como la práctica fueron muy bien presentadas por los docentes, y resultó de gran ayuda al aportar diferentes software con los que realizar el cálculo del tamaño muestral. Esta valoración positiva se recogió en la encuesta de satisfacción que se remitió a las personas asistentes y en la que se pedía que valorasen de 0 a 10 la adecuación de contenidos, el método de enseñanza, el interés de la materia, el profesorado, el cumplimiento de expectativas y la utilidad en el puesto de trabajo. Las valoraciones más altas se dieron en profesorado y en interés por la materia.

Como limitación, hemos de señalar la escasa presencia de participantes provenientes de economía y empresa, ciencias técnicas y arte y humanidades. Aunque la valoración general de la variabilidad en la participación es positiva, creemos que está altamente sesgada por nuestro entorno de trabajo y colaboraciones. En definitiva, nuestra capacidad de organización y difusión está limitada a un entorno concreto y, en posteriores ocasiones, deberíamos hacer un mayor esfuerzo por llegar más allá del mismo.

A modo de conclusión, la jornada puso de manifiesto el importante papel que juega la Estadística en la investigación de las diferentes ramas de conocimiento. Sin duda, podemos afirmar que la Estadística es necesaria en la investigación en general, y que además, al mismo tiempo, esa necesidad impulsará nuevas líneas de investigación estadística de cara a resolver los problemas planteados en otras áreas. A lo largo de la jornada, las personas participantes tuvieron ocasión de conocer las diferentes aplicaciones de la Estadística en diversas áreas y analizar su posible aplicabilidad en otros ámbitos. En nuestra opinión, la jornada fue un éxito y queremos agradecer a todas las personas que contribuyeron a ello, de manera especial a ponentes, participantes y patrocinadores. Nuestro deseo es que el éxito de esta jornada sirva de motivación para la organización de nuevos eventos de este tipo, y también, para que los diversos organismos universitarios tomen conciencia de la necesidad de este tipo de actividades y se impliquen de manera activa en la organización de las mismas.

Agradecimientos

Nuestro agradecimiento a las entidades que de una forma u otra han participado en la organización de este evento y han patrocinado el mismo, a saber,

la Facultad de Ciencia y Tecnología, el Departamento de Matemática Aplicada y Estadística e Investigación Operativa, el Grupo de Investigación en Bioestadística *Biostat* (IT620-13), la UFI en Matemáticas y Aplicaciones (UFI11/52) y la Sociedad Española de Biometría. Nuestro especial agradecimiento al personal del decanato, del servicio de multimedia y de la conserjería de la Facultad de Ciencia y Tecnología que nos facilitaron enormemente las labores técnicas. Finalmente, nuestro más sincero agradecimiento a Irati Murua y Josu Najera, estudiantes de grado y posgrado de Matemáticas, que trabajaron muchísimo en el desarrollo de la jornada.

Referencias

- [1] Abraira, V., Cadarso-Suárez, C., Gómez, G., Martín, A. y Pita, S. (2001). Mesa Redonda: La Estadística en la Investigación Médica. *Qüestió*, **25**, CLXXXII, 121-156.
- [2] Brown, E. N., and Kass, R. E. (2009). What Is Statistics?. *The American Statistician*, **63**, 105-110.
- [3] DeMets, D. L., Stormo, G., Boehnke, M., Louis, T.A., Taylor, J., y Dixon, D. (2006). Training of the Next Generation of Biostatisticians: A Call to Action in the U.S. *Statistics in Medicine*, **25**, 3415-3429.
- [4] Gordaliza, A. (2013) The time of statisticians has finally come. *BEIO*, **29(3)**, 229-245.
- [5] Lorenzo-Arribas, A., Martínez-Silva, I., Gómez-Mateu, M., Pérez-Álvarez, N., Perpignan-Fabuel, H., y Valero-Coppin, O. (2013) Young Biostatisticians in Spain: career or race?. *BEIO*, **29(3)**, 266-282.
- [6] Molenberghs, G. (2005). Biometry, biometrics, biostatistics, bioinformatics,..., bio-X. *Biometrics*, **61**, 1-9.
- [7] Mudelsee, M. (2010). *Climate time series analysis: classical statistical and bootstrap methods*, Springer, Alemania.
- [8] Polanco Martínez, J. M. (2011). *Aplicación de técnicas estadísticas en el estudio de fenómenos ambientales y ecosistémicos*. Tesis doctoral, servicio Ed. de la Universidad del País Vasco, <http://www.ehu.es/argitalpenak/images/stories/tesis/Ciencias/8120PolancoMartinezTH.pdf>
- [9] Schulz, M. y Mudelsee, M. (2002): REDFIT: estimating red-noise spectra directly from unevenly spaced paleoclimatic time series. *Computers and Geosciences*, **28**, 421-426.
- [10] Von Storch, H. y Zwiers, F. Z. (1999). *Statistical analysis in climate research*, Cambridge University Press, UK.

- [11] Zelen, M. (2006). Biostatisticians, Biostatistical Science and the Future. *Statistics in Medicine*, **25**, 3409-3414.

Acerca de las autoras



Inmaculada Arostegui Madariaga es profesora agregada en el Departamento de Matemática Aplicada, Estadística e Investigación Operativa de la Universidad del País Vasco UPV/EHU. Es además asesora estadística del Hospital Galdakao–Usansolo, desde 1997. Licenciada y doctora en Ciencias Matemáticas, sus líneas de investigación actuales se centran en la bioestadística aplicada a la epidemiología clínica y la investigación de servicios sanitarios, y más concretamente en la medición y análisis de resultados en salud.

Ha publicado artículos metodológicos y aplicados en revistas de impacto e imparte docencia en programas de grado y posgrado de Ciencias Experimentales y Medicina. Lidera el grupo de Bioestadística de la UPV/EHU (*Biostat*) y tiene colaboraciones científicas con otros grupos de investigación, tanto del ámbito universitario como de otros organismos y empresas.



Irantzu Barrio Beraza es profesora interina en el Departamento de Matemática Aplicada, Estadística e Investigación Operativa de la Universidad del País Vasco UPV/EHU. Su principal línea de investigación es el desarrollo de modelos predictivos para la aplicación en la práctica clínica, y más concretamente la categorización óptima de variables continuas en modelos de regresión logística y supervivencia. Adicionalmente, colabora como asesora bioestadística en la unidad de investigación del Hospital de Galdakao–Usansolo.



Ma José Legarreta Olabarrieta es investigadora contratada del Departamento de Matemática Aplicada, Estadística e Investigación Operativa de la Universidad del País Vasco UPV/EHU. Está adscrita al Grupo de Investigación en Bioestadística (*Biostatit*) y su actividad se centra en el desarrollo de aplicaciones informáticas a partir de modelos predictivos para dar soporte a la toma de decisiones en la práctica clínica.



Arantza Urkaregi Etxepare es profesora agregada en el Departamento de Matemática Aplicada, Estadística e Investigación Operativa de la Universidad del País Vasco UPV/EHU. Es además asesora estadística del Hospital Universitario Cruces, desde 1994. Su línea de investigación se centra en técnicas estadísticas para el análisis de encuestas y construcción de scores de evaluación de programas y servicios sanitarios a partir de las respuestas a encuestas. Forma parte del Grupo de Investigación Bioestadística de la UPV/EHU (*Biostatit*) y colabora con otros grupos de

investigación tanto universitarios como de otros organismos y empresas. Miembro del Grupo de Desarrollo Estadístico creado por el EUSTAT-Instituto Vasco de Estadística para la elaboración del Plan Vasco de Estadística 2013-2016.