

TENDIENDO LAZOS ENTRE JÓVENES INVESTIGADORES Y EXPERTOS EN ESTADÍSTICA: UNA EXPERIENCIA IBEROAMERICANA DE ANÁLISIS DE DATOS COLABORATIVO

Monica Balzarini¹, Cecilia Bruno¹, Teresa Boca², Franca Giannini Kurina¹, Pablo Paccioretti¹,
Adriana Pérez², y Silvia Sühling³

¹Universidad Nacional de Córdoba

²Universidad de Buenos Aires

³Universidad Nacional de Salta

monica.balzarini@unc.edu.ar

Para fortalecer la formación en estadística en el nivel de posgrado, propusimos webinarios abiertos de análisis de datos colaborativo entre tesis doctorales y especialistas en biometría de Iberoamérica. Los webinarios fueron promovidos por la Sociedad Internacional de Biometría. El objetivo de este trabajo es evaluar estos webinarios llevados a cabo con modalidad de taller y abordaje interdisciplinario como estrategia de enseñanza/aprendizaje de la planificación e implementación de análisis estadístico de datos en biociencias. En cada webinar, tres tesis presentaron públicamente su problema de investigación a biometristas quienes promovieron el debate público sobre ventajas/limitaciones de alternativas analíticas para el tratamiento de los datos provenientes de investigaciones doctorales, tutoraron la implementación del análisis y finalmente la comunicación efectiva de los resultados estadísticos. El modelo propuesto promovió la adquisición de conceptos y métodos del análisis de datos desde la práctica y la reflexión participativa.

CONTEXTO

Las tesis de nivel de doctorado en ciencias de la vida usualmente generan bases de datos multidimensionales con complejas estructuras de correlación que dificultan la modelación estadística e incluso la aplicación de métodos exploratorios. La formación en estadística de quienes se están especializando en una disciplina científica distinta a la del análisis de datos, usualmente no alcanza a cubrir las necesidades reales de conocimiento y de manejo solvente de nuevas herramientas de análisis que se requieren para un análisis eficiente de esos datos, es decir, para obtener mayor cantidad de información por unidad de recurso asignado a la investigación. Esto ha llevado a la necesidad de implementar distintas estrategias en la enseñanza de estadística en las biociencias (Weissgerber et al., 2016).

Nuestra propuesta se enfocó en enseñar y valorizar el análisis de datos a través de la modalidad de taller con abordaje interdisciplinario para la planificación e implementación de estrategias de análisis estadístico de datos de biociencias, partiendo de la hipótesis de que la interacción de los investigadores con más de un especialista en estadística permitirá ampliar el espectro del espacio de soluciones y fortalecer las capacidades analíticas de jóvenes investigadores.

Pretendimos promover la importancia de trabajar con datos desde distintas miradas, entender que existen diferentes posibilidades de análisis y que es necesario tener formación en los métodos y modelos estadísticos específicos que se utilizan en el tópico que se investiga. La práctica de aprendizaje basado en proyectos es un método de enseñanza-aprendizaje conocido para desarrollar experiencia (Takemura et al., 2018), en la que los estudiantes llevan a cabo proyectos bajo condiciones reales para adquirir conocimientos y desarrollar habilidades. La innovación aquí es que esa práctica se realiza con la participación de quienes investigan en biociencia con más de una persona experta para promover distintas miradas, y que la construcción de conocimiento se da a partir del debate sobre las potenciales estrategias analíticas que puedan generar mayor valor desde los datos para el problema que se investiga.

La finalidad de esta propuesta educativa es mejorar las competencias en análisis de datos a través de la práctica del análisis colaborativo, la comunicación de problemas científicos en términos de datos e hipótesis estadísticas y la práctica del análisis con métodos y modelos estadísticos de actualidad mediante la asistencia y supervisión de personas expertas en ciencia de datos.

MÉTODOS: DISEÑO DE LOS TALLERES DE ANÁLISIS ESTADÍSTICO COLABORATIVO

Los talleres se llevaron a cabo como webinarios en modalidad virtual y con instancia de trabajo off-line (Figura 1). Los webinarios se enmarcaron en alguna de tres áreas temáticas de investigación en

biociencias donde se recolectan distintos tipos de datos: Ecología-Ambiente, Epidemiología-Ciencias de la Salud y Ciencias Ómicas. Se organizó un webinar de Análisis Estadístico Colaborativo (CADA, del inglés Collaborative Approach for Data Analysis) para cada área.

Cada uno consistió en dos encuentros virtuales sincrónicos abiertos a la comunidad académico-científica y un encuentro *off-line* entre tesista y biometristas. El primer encuentro virtual fue destinado a la presentación del caso de investigación y de los datos por parte del investigador/a joven y posterior discusión y propuesta de estrategias de análisis por parte de especialistas; el segundo, a la comunicación de los resultados obtenidos y la discusión de los alcances y limitaciones de las propuestas de análisis. Luego del primer encuentro, los/las doctorandos/as resolvieron los problemas con la colaboración de las personas expertas, tanto a nivel de implementación de rutinas de análisis como a nivel de la comunicación efectiva de los resultados estadísticos; esta actividad fue realizada fuera del horario del webinar.

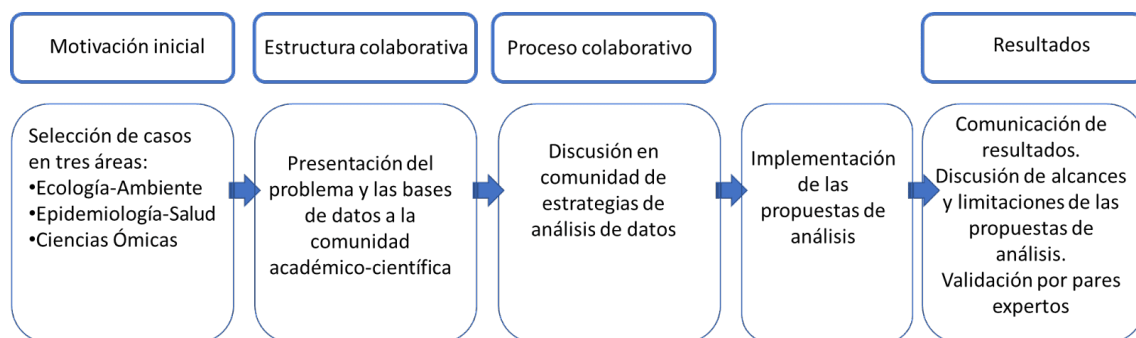


Figura 1. Diseño de la propuesta de análisis colaborativo de datos

Para el desarrollo de los webinarios se invitó a participar como tutores para el análisis de datos a analistas especialistas en Biometría con experiencia en consultoría de distintas regiones de la Sociedad Internacional de Biometría (IBS). Participaron analistas de Argentina, Brasil, América Central y Caribe, Ecuador, Chile, y España. Cada joven investigador/a trabajó con dos especialistas atentos a sugerir/enseñar/liderar el análisis de los datos y la interpretación de los resultados obtenidos.

Se realizó una convocatoria para estudiantes de posgrado que estuvieran realizando investigación científica en alguna de esas tres áreas. Los postulantes debían estar cursando un posgrado, explicar el objetivo a responder mediante el análisis estadístico de los datos y su contexto, explicitar cómo obtuvo los datos y mostrar parte de la base de datos afín al objetivo del análisis estadístico. Además, debían contar con la conformidad explícita del trabajo público por parte del director/a de tesis. Debido a que son datos en proceso de investigación para posgrado, los mismos no son públicos.

La invitación se distribuyó a través de canales formales a Universidades y Sociedad de Estadística de Argentina y de otros países de Iberoamérica, y a través de redes sociales y de la página del Grupo Argentino de Bioestadística (en www.gab.com.ar).

Un comité internacional, conformado por biometristas de la IBS de las regiones participantes, seleccionó tres casos de investigación entre las postulaciones para cada área temática, atendiendo a la diversidad geográfica, a la disponibilidad de datos y a la diversidad de herramientas de análisis que podrían utilizarse.

Por otro lado, una vez definidos los casos a abordar en cada taller y el cronograma correspondiente, se realizó una convocatoria a la comunidad científico-académica relacionada con la Bioestadística para participar como oyentes. Se utilizaron los mismos canales de difusión y se requirió la inscripción previa a los interesados en presenciar los webinarios.

RESULTADOS

La convocatoria a los webinarios tuvo una rápida aceptación que dio como resultado casi un centenar de solicitudes de participación de jóvenes investigadores en Biociencias, la mayoría (84%) provenientes de Argentina y el resto de Chile, Ecuador, Colombia, Perú, y Nicaragua. Se seleccionaron siete investigaciones realizadas en Argentina, una en Chile, una en Ecuador, y una en Nicaragua. En

función de la cantidad de postulaciones se seleccionaron cuatro casos para el área de Ecología y Ambiente y tres para las otras dos áreas.

Los problemas abordados por cada estudio y los objetivos de análisis fueron presentados por cada estudiante de posgrado usando la plataforma web en el primer encuentro. Las personas expertas en Biometría asignados a cada caso presentaron ideas sobre cómo abordar el análisis y el porqué de la sugerencia. En todos los casos se sugirió más de una alternativa de análisis. Todas las estrategias de análisis incluyeron una etapa de depuración, preprocesamiento y visualización de los datos disponibles para su posterior procesamiento estadístico. En el webinar sobre Epidemiología-Ciencias de la Salud se trabajó con métodos multivariados para la interpretación de la interdependencia entre las múltiples variables colectadas sobre cada individuo y la construcción de clasificadores asociados a estados de interés. En el webinar de Ciencias Ómicas se abordó la construcción del dato para análisis estadístico a partir de la información molecular recabada en el laboratorio y la posterior comparación y selección de modelos para identificar variantes genómicas asociadas con variabilidad en la respuesta de interés. En el webinar sobre Ecología-Ambiente, se profundizó en la construcción de modelos mixtos para datos no-normales de variables identificadas como respuestas de procesos multifactoriales, en la evaluación de los modelos y en la comparación de resultados desde modelos alternativos. En todos los webinars fue importante la discusión sobre cómo presentar en tablas y/o gráficos los resultados de los análisis estadísticos para su comunicación efectiva. Los talleres incluyeron un periodo de trabajo conjunto entre las personas expositoras de los problemas de análisis y las que se desarrollaron como expertas, espacio propicio para enseñar/aprender herramientas de softwares estadísticos, así como nuevas implementaciones de herramientas de análisis.

Se registraron más de 200 participantes de varios países iberoamericanos conectados durante las transmisiones en vivo de estos webinars (Tabla 1). Se brindó la posibilidad de realizar sugerencias sobre alternativas de análisis a los oyentes a través del chat durante la conferencia, la que fue resumida y entregada a los expositores. La gran afluencia de participantes como oyentes amplió de manera exponencial la llegada de sugerencias vertidas por expertos dando respuesta a problemas de las biociencias respecto a la mejor estrategia de análisis. Los ejes temáticos seleccionados permitieron abordar entre los tres webinars un gran abanico de problemas y estructuras de bases de datos. Los webinars como estrategia de enseñanza y generación de lazos permitieron potenciar el aprendizaje no solo de las personas que presentaron y avanzaron en análisis de datos de su tesis sino de todas aquellas que participaron como oyentes y que pudieron capturar sugerencias de análisis que se podrían extrapolar a otros problemas del área disciplinar. Así, el trabajo en taller de análisis de datos promovió que investigadores en formación adquieran habilidades y destrezas para el análisis crítico de datos provenientes de estudios experimentales u observacionales, y conformó un espacio público propicio para dar a conocer nuevos desarrollos y aplicaciones de la biometría.

Tabla 1. Participación de los oyentes en cada área temática

Taller	Inscriptos	Duración del Webinar (min)	Tiempo de conexión promedio (min)
Ecología y ambiente	139	184	130
Epidemiología y Salud	71	120	88
Datos ómicos	57	130	85

Finalmente, en la segunda sesión pública se expusieron los resultados que fueron debatidos con participación de personas del público. Si bien se trabajaron problemas específicos se destacaron las estructuras comunes y se puso en evidencia las ventajas y limitaciones de análisis alternativos en relación con la generación de conclusiones sobre el problema investigado. Los videos de los encuentros quedaron disponibles en la página web del Grupo Argentino de Biometría (<http://www.gab.com.ar>). Los webinars promovieron la conceptualización del análisis de datos desde la práctica y la reflexión participativa no solo entre los jóvenes investigadores seleccionados si no entre los oyentes que pudieron

comprender la estructura de las bases de datos a ser analizadas, la recepción de las sugerencias vertidas por expertos, la discusión del público en general que se podía intercambiar a través de mensajes durante la sesión virtual y la extrapolación a su propia base de datos para abordar sus análisis. Esta interacción entre integrantes de la comunidad científica permitió difundir la actividad del bioestadístico/a y valorizar el análisis de datos a través del trabajo interdisciplinario. A pesar de las dificultades conocidas en la interacción entre estadísticos y jóvenes científicos y dada la diversidad y complejidad de diseños experimentales y temas de investigación, consideramos que la experiencia constituyó un valioso puente entre estudiantes y biometristas. Quienes participaron con sus problemas de tesis manifestaron que los webinarios fueron muy enriquecedores ya que permitieron conocer nuevas herramientas y programas de análisis y cuáles son las maneras apropiadas de presentar los resultados. La colaboración de expertos/as en ciencia de datos jugó un rol clave en el desarrollo y ejecución de la propuesta. La propuesta educativa permitió no sólo abordar conceptos y métodos para profundizar el análisis de datos de investigaciones en biociencias, sino también difundir la actividad del bioestadístico/a. El aprendizaje basado en proyectos es un método que promueve el aprendizaje significativo, a través de la resolución de problemas reales usualmente propuestos por el cuerpo docente. En la presente propuesta esta modalidad resultó aún más motivadora ya que los participantes trabajaron con datos propios de su investigación.

CONCLUSIÓN

La era digital y la comunicación virtual abrió nuevas posibilidades para innovar sobre la enseñanza/aprendizaje del análisis de datos a nivel de posgrado. Las reuniones virtuales a través de plataforma web, permitieron la participación de investigadores jóvenes y analistas con experiencia en aplicaciones del análisis de datos de diferentes regiones, aumentando no sólo la diversidad de los problemas sino también la participación de más personas para implementar validaciones cruzadas y consensos de estrategias de análisis.

Los talleres de análisis colaborativo de datos promovieron la articulación entre estudiantes de posgrado y especialistas en consultoría estadística y ciencia de datos. Por otro lado, permitió generar una red de analistas e investigadores/as jóvenes con proyección de continuar la cooperación.

AGRADECIMIENTOS

Las actividades aquí presentadas corresponden al Proyecto “Ibero-American Statistical Analysis Board (ISAB)” y fueron subsidiadas por la Sociedad Internacional de Biometría (IBS), a través de su Programa de Financiamiento de Redes y Actividades Interregionales.

Queremos agradecer a nuestros colegas de la IBS que participaron activamente de la convocatoria, tanto para la difusión entre estudiantes de posgrado como para la propuesta de especialistas biometristas: Pablo Reeb, Fernando Casanoves y Pablo Demetrio (Argentina), Clarice García Borges Demetrio y Paulo Canas (Brasil), Raúl Macchiavelli (Puerto Rico), Danilo Alvares (Chile), Omar Ruiz Barloza (Ecuador), Lupe Gómez Melis, Klaus Langohr, Nuria Porta, Pere Puig y Anabel Forte (España). También a los biometristas que actuaron como tutores: Nora Abbiati, Mónica Balzarini, Teresa Boca, Cecilia Bruno, Gerardo Cueto, Franca Giannini-Kurina, Susana Fillipini, Myriam Núñez, Guillermo Pratta, Adriana Pérez y Ezequiel Rossi, (Argentina), Augusto Franco García (Brasil), Purificación Galindo Villardón, Virgilio Gómez Rubio, Alex Sánchez (España), Danilo Alvares (Chile), Sergio Vilchez (Costa Rica), Raúl Macchiavelli (Puerto Rico). A la Universidad Nacional de Salta, Universidad de Buenos Aires, Universidad Nacional de Córdoba y al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) por el soporte brindado.

REFERENCES

- Takemura, A., Izumi, S., Saito, K., Himeno, T., Matsui, H., & Date, H. (2018). Shiga University model of data science education. *Proceedings of the Institute of Statistical Mathematics*, 66(1), 63–78.
- Weissgerber, T. L., Garovic, V. D., Milin-Lazovic, J. S., Winham, S. J., Obradovic, Z., Trzeciakowski, J. P., & Milic, N. M. (2016). Reinventing biostatistics education for basic scientists. *PLoS Biology*, 14(6), Article e1002430. <https://doi.org/10.1371/journal.pbio.1002430>