

Bioestadística y bioinformática: fundamentales para analizar datos

Los beneficios de la bioestadística y la bioinformática pueden llegar a ser indirectos, sin embargo, son de gran ayuda para la población.

Pablo R. Martínez

22 de mayo de 2023

Se necesitaron 13 años y 3 mil millones de dólares para que el Proyecto Genoma Humano lograra secuenciar el genoma humano. Lo que antes tomaba meses, ahora se puede realizar en un día y con un costo de unos pocos miles de dólares gracias a herramientas que nos permiten analizar y desmenuzar millones de datos.

Dos de esas herramientas, la bioestadística y la bioinformática, son también dos protagonistas del curso "Bioestadística y Bioinformática para Ciencias Biomédicas" que se imparte en el Centro de Ciencias de la Complejidad (C3), UNAM, por las investigadoras de la Red de Apoyo a la Investigación (RAI), de la UNAM, Nancy R. Mejía Domínguez y Georgina Hernández Montes.



Nancy Raquel Mejía imparte la primera clase del curso "Bioestadística y Bioinformática para Ciencias Biomédicas".
Foto de Pablo R. Martínez

"La bioestadística es una disciplina que estudia cómo se colectan, organizan y analizan datos y las conclusiones a las que se puede llegar con ellos", dice Mejía Domínguez, quien actualmente trabaja en análisis estadísticos aplicados a la biomedicina y enfermedades no transmisibles.

Por su lado, "la bioinformática es una disciplina que hace uso de bases de datos, modelos estadísticos, herramientas computacionales y de conocimientos biológicos, integrándolos para poder responder a las preguntas biológicas moleculares", agrega Hernández Montes, cuya principal línea de investigación son los análisis bioinformáticos enfocados a la biomedicina.

Así que mientras que la bioestadística ayuda a interpretar y entender los datos que se obtienen en estudios médicos y a tomar decisiones informadas basadas en esos datos, la bioinformática utiliza la tecnología y la informática para entender la biología de los organismos y sus procesos biológicos, desde el nivel molecular hasta el nivel de poblaciones completas.

Combatiendo el COVID-19

Aunque todos podemos reconocer los beneficios de las vacunas para combatir COVID-19, el control y conocimiento de la enfermedad no habría sido posible sin la bioestadística ni la bioinformática.

En el caso de la bioinformática, “una de las cosas que se nos permitió fue hacer los análisis filogenéticos, con programas bioinformáticos basados en bioestadística. Se hizo una secuenciación de los virus, encontrando variantes en las personas que podía determinar su sensibilidad, es decir, como se vio en la pandemia, había personas que enfermaban más grave que otras y eso tenía que ver con su información genética, por lo que la bioinformática se encargó de estudiar cuáles eran esas características”, mencionó Georgina Hernández.

Gracias a que se consiguió secuenciar por primera vez el genoma completo del virus SARS-CoV-2, por Zhang Yongzhen en Shangai a principios del año 2020, se identificaron variantes del virus en Gran Bretaña, Sudáfrica, Brasil y en la India. Además, se logró saber cómo es que atacan al sistema inmunitario y, a partir de ese avance, se comenzó a pensar en cómo inhibir su replicación.

Por su parte, “con bioestadística se realizó el análisis de los efectos de edad y el sexo sobre la mortalidad o morbilidad asociada al COVID-19 y también se encargó de ver la eficacia de las vacunas, pues antes de que la vacuna salga al mercado tiene que hacerse diversos análisis para ver si realmente funciona o no en la población”, agregó Nancy Mejía.

Diversos grupos utilizaron la bioestadística para identificar si había alguna relación entre los estratos de edad y sexo con la mortalidad o morbilidad que se asociaba al COVID-19, es decir, si la mortalidad o la morbilidad afectaba más a hombres o a mujeres y qué grupos de edades eran más vulnerables ante este virus.

Otros beneficios para la investigación y la medicina

Puede que no veamos beneficios directos de la bioestadística y la bioinformática, pero indirectamente nos permiten tener importantes avances en la investigación y también en la salud pública.

Ronald A. Fisher, estadístico y biólogo británico, fallecido en 1962, es considerado uno de los fundadores de la bioestadística moderna, pues hizo importantes contribuciones a esta disciplina a inicios del siglo XX, las cuales se utilizan de manera práctica en la investigación biomédica y la salud pública. Algunas de ellas, por ejemplo, benefician a grupos de investigación cuando quieren tener resultados más confiables, pues permite perfeccionar técnicas para el diseño óptimo de experimentos; otra herramienta de la bioestadística es el análisis de varianza (ANOVA), que es una técnica que se utiliza para comparar las medias de tres o más grupos de datos y determinar si hay diferencias significativas entre ellos.

Además, la bioinformática ha permitido identificar genes asociados a enfermedades complejas, como la diabetes y el cáncer, lo que ha llevado al desarrollo de nuevas terapias y tratamientos personalizados para los pacientes.

La bioestadística, por su parte, ha contribuido al desarrollo de pruebas diagnósticas más precisas para enfermedades infecciosas, como el VIH y la tuberculosis, logrando así una detección temprana y un tratamiento efectivo de estas enfermedades, lo que ha salvado muchas vidas.

Precisamente, la bioinformática surgió de la necesidad de manejar y analizar grandes cantidades de datos biológicos. Este término fue utilizado por primera vez en 1978 por el matemático y biólogo estadounidense Walter Fitch, quien utilizó la palabra para describir el estudio de la información procesada por los sistemas

biológicos. Fitch es reconocido como el padre de la bioinformática, pues fue uno de los primeros investigadores en aplicar métodos matemáticos y estadísticos al análisis filogenético, que es la forma en que los científicos reconstruyen la historia evolutiva de los seres vivos utilizando datos moleculares, morfológicos y/o de comportamiento.

Para las investigadoras, cualquier estudiante que quiera adentrarse en la biología o la medicina, requerirán el conocimiento indispensable de estas disciplinas, especialmente, en una época en la que atravesaremos importantes desafíos para la salud por la aparición de enfermedades emergentes.

“Estas disciplinas son muy importantes, sin estas herramientas no habría manera de analizar los datos. Actualmente, los datos biológicos ocupan el tercer o cuarto lugar de generación de datos masivos en la red o *big data*, entonces no habría manera de analizar los miles y miles de genomas que se encuentran en las bases de datos”, dijo Hernández Montes.

“Es un pilar para que se desarrolle la biomedicina. Si no utilizáramos estas herramientas, no habría forma de llevar a cabo, por ejemplo, el desarrollo de vacunas”, concluyó Mejía Domínguez.

Ligas de interés

Perfil de Nancy Raquel Mejía Domínguez:

<https://sites.google.com/view/biostatisticsrai-unam/nancy-r-mej%C3%ADa-dom%C3%ADnguez>

<https://www.researchgate.net/profile/Nancy-Mejia-4>

Perfil de Georgina Hernández Montes:

<https://www.researchgate.net/profile/Georgina-Hernandez-Montes>