

ARTÍCULO DE REVISIÓN

Métodos gráficos en la investigación biomédica de causalidad Graphic methods in causal biomedical research

Nelsa María Sagaró-del-Campo¹, Larisa Zamora-Matamoros²

¹Universidad de Ciencias Médicas de Santiago de Cuba, Facultad de Medicina No. 1, Santiago de Cuba. ²Universidad de Oriente, Santiago de Cuba, Cuba. **Correspondencia a:** Nelsa María Sagaró-del-Campo, correo electrónico: nelsa@infomed.sld.cu

Recibido: 19 de abril de 2019

Aprobado: 7 de junio de 2019

RESUMEN

Fundamento: los gráficos estadísticos pueden servir para el análisis exploratorio de los datos, siendo indispensables en el estudio de relaciones multivariadas.

Objetivo: actualizar aspectos referentes al empleo de los métodos gráficos en el estudio de la causalidad en investigaciones de las ciencias biomédicas.

Métodos: se realizó una revisión bibliográfica utilizando servicios especializados disponibles en internet: Pubmed/Medline, SciELO, SCOPUS, Springer, Web of Science, EBSCOhost, Google; entre enero y marzo de 2019. Se emplearon como descriptores, en idiomas inglés, francés, portugués y español: técnicas estadísticas, gráficos estadísticos, diagrama causal, grafo implicativo, relaciones entre variables, análisis estadístico implicativo. Se incluye la experiencia de las autoras.

Resultados: se presentan los métodos gráficos más empleados, donde se destacan los que permiten presentar la información, como los diagramas causales que visualizan la relación entre múltiples variables y los que permiten la exploración de los datos y las relaciones multivariadas, para dirigir el análisis posterior. Se argumenta la utilidad del grafo implicativo para identificar los factores que influyen sobre un desenlace, permitiendo distinguir cuando se trata de un factor de riesgo o protector y cuál es la magnitud del efecto, a través de la intensidad implicativa.

Conclusiones: se evidenció la importancia del empleo de los métodos gráficos en el estudio de la causalidad. Se destaca la superioridad del grafo implicativo.

Palabras clave: TÉCNICAS ESTADÍSTICAS; GRÁFICOS ESTADÍSTICOS; DIAGRAMA CAUSAL; GRAFO IMPLICATIVO; RELACIONES ENTRE VARIABLES; ANÁLISIS ESTADÍSTICO IMPLICATIVO.

Descriptores: ESTADÍSTICA COMO ASUNTO; ANÁLISIS ESTADÍSTICO; INVESTIGACIÓN BIOMÉDICA, MÉTODOS DE ANÁLISIS.

ABSTRACT

Background: the statistical graphs can be used for the exploratory data analysis, being indispensable in the study of multivariate relations.

Objective: to update on the aspects related to the use of graphic methods in the study of causality in biomedical sciences.

Methods: a bibliographic review was conducted using the specialized services available in the Internet: Pubmed/Medline, SciELO, SCOPUS, Springer, Web of Science, EBSCOhost and Google, from January through March 2019. The following descriptors in English, French, Portuguese and Spanish were used: statistical techniques, statistical graphs, causal diagrams, implicative graph, variable relations and implicative statistical analysis. The professional experience of the authors was included.

Results: the mostly used graphic methods are presented, highlighting those that allow the presentation of the information such as the causal diagrams that visualize the relationship between multiple variables, as well as those that allow the exploration of data and multivariate relations to direct the subsequent analysis. The study shows the usefulness of the implicative graph to identify the factors that have an influence on an ending, making it possible to determine whether it is a risk or a protective factor, as well as the magnitude of the effect, by means of implicative intensity.

Citar como: Sagaró-del-Campo NM, Zamora-Matamoros L. Métodos gráficos en la investigación biomédica de causalidad. Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta. 2019; 44(4). Disponible en: <http://revzoilomarinellosld.cu/index.php/zmv/article/view/1846>.



Conclusions: the study showed the importance of the use of graphic methods in the study of causality. The superiority of the implicative graph is highlighted.

Key words: STATISTICAL TECHNIQUES; STATISTICAL GRAPHS; CAUSAL DIAGRAM; IMPLICATIVE GRAPH; RELATIONS BETWEEN VARIABLES; IMPLICATIVE STATISTICAL ANALYSIS.

Descriptors: STATISTICS AS TOPIC; STATISTICAL ANALYSIS; BIOMEDICAL RESEARCH; ANALYTICAL METHODS.

INTRODUCCIÓN

El propósito de un gráfico estadístico es ayudar a la comprensión y comunicación de la evidencia aportada por los datos. Su utilidad es doble, ya que sirven tanto para presentar la información elaborada previamente y dispuesta en una tabla, como para el análisis exploratorio de los datos, siendo en ocasiones un paso indispensable en el procesamiento de la información. Por ejemplo, para identificar la posible correlación entre dos variables cuantitativas, el primer paso es construir un diagrama de dispersión, cuya observación permite continuar el análisis hacia una u otra vertiente.

En el caso univariado y bivariado, los gráficos de presentación sirven para percibir más fácilmente las características esenciales de una distribución de frecuencias, comparar una distribución con otra, evidenciar de manera objetiva las relaciones y evolución de fenómenos poco aparentes en el material tabulado; pero nunca podrán sustituir las tablas, donde se plasman los datos exactos y de la cual son un complemento o ayuda visual. En cambio, las relaciones multivariadas no pueden ser presentadas en tablas y menos cuando se trata de un gran número de variables, donde hasta el gráfico resulta complejo de entender. De ahí la necesidad e importancia de conocer los métodos gráficos para el estudio de la causalidad, donde se analizan relaciones multivariadas. En este entorno se ubica el problema que motivó esta investigación bibliográfica.

En respuesta a ello, este trabajo tiene como objetivo proporcionar una síntesis de los métodos gráficos que pueden ser empleados en el estudio de la causalidad en las ciencias biomédicas, tanto los diagramas causales, que permiten visualizar la relación entre múltiples variables, como los gráficos de exploración de las relaciones multivariadas, que permiten dirigir el análisis posterior.

Así, se realizó una exhaustiva búsqueda bibliográfica en bases de datos biomédicas disponibles en Internet, entre enero y marzo de 2019, bases y otros servicios tales como: Pubmed/Medline, SciELO, SCOPUS, Springer, Web of Science, EBSCOhost y Google. En la estrategia de búsqueda se emplearon como descriptores los consignados en las palabras clave de este artículo: técnicas estadísticas, gráficos estadísticos, diagrama causal, grafo implicativo, relaciones entre variables, análisis estadístico implicativo. Las búsquedas se realizaron en idiomas inglés, francés, portugués y español, combinados con los operadores lógicos, seleccionando la literatura más apropiada por su carácter científico y actualizado e identificando en ellas regularidades y diferencias que permitieron la elaboración de un

material resumen de estas técnicas, donde se plasmó, además, la experiencia de las autoras.

DESARROLLO

Se presentan en este material los diagramas causales, que representan la interrelación entre múltiples variables que, según el algoritmo empleado para obtener dichas relaciones, adquieren diferentes denominaciones: un conjunto de métodos gráficos que apoyan las técnicas multivariadas empleadas en los estudios de causalidad, los gráficos de exploración empleados previo a estas técnicas multivariadas y los gráficos espaciales. Los diferentes tipos de gráficos se podrán visualizar consultando la literatura citada.

Métodos gráficos para el análisis de las relaciones entre múltiples variables

Los diagramas que denotan causalidad se basan en la teoría de grafos. Los mismos permiten visualizar un inmenso número de nodos entrelazados, representando las relaciones entre el conjunto de variables. Estos diagramas han sido abordados en inteligencia artificial, empleando muchos algoritmos para su construcción. ⁽¹⁾ También, existen diversas aplicaciones y entornos web que facilitan este propósito. ⁽²⁾

El grafo acíclico dirigido

El diagrama causal más empleado es el grafo acíclico dirigido (DAG del inglés *Directed Acyclic Graph*), el cual se considera una herramienta estadística poderosa. ^(2,3) Estos gráficos resultan de gran utilidad, ya que plantean posibles rutas de asociación entre causas y efecto, así como otras rutas alternativas generadoras de sesgos, por lo que facilitan también el control de estos, ⁽⁴⁻⁷⁾ además, guían al investigador en cuanto a cuáles variables deben incluirse en el análisis.

Los DAG comenzaron a emplearse en epidemiología por James Robins en 1987. ⁽⁸⁾ En 1993, Judea Pearl formalizó este diagrama, basado en las probabilidades condicionales, tomando la denominación de redes bayesianas. ⁽⁹⁾ Estos gráficos permiten evaluar la existencia de sesgos de selección o de confusión. ⁽¹⁰⁻¹²⁾

Las categorías y elementos básicos a tener en cuenta en un DAG son:

- Nodo o vértice, que representa a una variable o conjunto de ellas.
- Camino dado por cualquier ruta, que conecte la exposición (X) al evento (Y), independiente de la dirección de la flecha.
- El efecto total de X sobre Y puede ser estimado a partir de las distribuciones observadas del resto de las variables.

- La relación entre dos variables o dos conjuntos de variables (X e Y) puede verse bloqueada (d-separada) por una tercera variable, o conjunto de ellas (Z) que se interpone en el camino, que enlaza ambas variables y esto puede ocurrir de tres formas. ^(5,11-15)

- Cadena o “camino causal dirigido”, si partiendo de la exposición mantiene su dirección hasta llegar al evento ($X \rightarrow Z \rightarrow Y$).

- División (*fork*), “camino de puerta trasera” o vía no causal: si la vía incluye alguna flecha apuntando en dirección opuesta. ($X \leftarrow Z \rightarrow Y$) es un camino de confusión, donde debiera ajustarse por Z, factor de confusión asociado con la exposición y el evento.

- División invertida, “camino de choque” o colisionador: si dos puntas de flecha “chocan” o se encuentran en la variable Z denominada “collider”; ($X \rightarrow Z \leftarrow Y$) indica que no es necesario ajustar, pues la vía naturalmente está bloqueada.

En el control de confusión se evalúa la existencia de vías no casuales o “caminos de puerta trasera” abiertos, que puedan transmitir una asociación espuria que sesgue la estimación de la asociación. Se buscará el ajuste por esa causa común que “cierra” todos los posibles caminos no causales entre la exposición y el evento.

Estos grafos también permiten valorar el cambio en la inferencia causal, cuando se emplean variables instrumentales, otro método que permite el control de los confusores. ⁽¹⁶⁾

Moreno-Betancur ⁽¹⁷⁾ expone una versión de este grafo para tratar los datos perdidos, el DAG canónico; lo cual teniendo en cuenta la frecuencia de datos perdidos en las investigaciones médicas, que trabajan con grandes bases de datos, resultan de gran interés.

Las redes bayesianas

Las redes bayesianas son un método estadístico para la representación del conocimiento incierto, que permite establecer razonamientos basados en las probabilidades condicionadas y el teorema de Bayes, constituyendo un modelo probabilístico asociado a un modelo gráfico, basado en la teoría de grafos. Estas redes no son más que un grafo formado por variables aleatorias, enlazadas por una distribución de probabilidades. ⁽¹⁸⁾ Para el cálculo de estas probabilidades existen diferentes algoritmos. ⁽¹⁹⁾ En estos grafos los nodos representan las variables y los arcos, o enlaces entre los nodos, los mecanismos causales o dependencias probabilísticas. Esta es una de las aplicaciones de la inteligencia artificial más difundida en la actualidad. El surgimiento de estos modelos permitió el estudio de problemas con cientos y hasta miles de variables, como los relacionados con la genética. ⁽²⁰⁾

Las redes neuronales artificiales

Las redes neuronales artificiales es otro método propio de la inteligencia artificial, que surgió en la década de los 50 y aún continúa en desarrollo. Estas redes constituyen un modelo matemático computacional, que simula las interconexiones de las neuronas del cerebro humano que intentan “aprender” de los datos que se le suministran. La

idea es que, dados unos parámetros, haya una forma de combinarlos para predecir un cierto resultado; por ejemplo, sabiendo los píxeles de una imagen habrá una forma de saber qué número está escrito. Estas redes se pueden representar, también, mediante grafos, donde los nodos son las neuronas y las conexiones equivalen a las sinapsis. Las mismas pueden adoptar diferentes arquitecturas, así mismo, existen múltiples algoritmos para su entrenamiento y activación. ⁽¹⁹⁾

Los árboles de decisión

Los árboles de decisión son modelos gráficos de predicción, utilizados en el ámbito de la inteligencia artificial, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema. El árbol tiene nodos internos, nodos de probabilidad, nodos hojas y arcos. Son muy empleados en el análisis de decisión clínica y en la evaluación económica en salud. Permiten la segmentación, la estratificación, la predicción, la reducción de datos, el filtrado de variables, la identificación de interacciones, la fusión de categorías y la discretización de variables continuas. ⁽²¹⁾

Los mapas cognitivos difusos

Los mapas cognitivos difusos constituyen una estructura de grafo difuso con retroalimentación, para representar causalidad. Esta es una técnica que combina herramientas teóricas de los mapas cognitivos, la lógica difusa, las redes neuronales, las redes semánticas, los sistemas expertos y los sistemas dinámicos no lineales. Estos mapas permiten describir la fortaleza de la relación mediante el empleo de valores borrosos en el intervalo [-1,1]. Los nodos son conceptos causales y pueden modelar eventos, acciones, valores, metas o procesos. ⁽²²⁾ Este tipo de modelo puede ser útil, por ejemplo, para evaluar competencias. ⁽²³⁾

El grafo implicativo

El grafo implicativo es un gráfico generado dentro del análisis estadístico implicativo, que muestra las relaciones causales entre el conjunto de variables. ⁽²⁴⁻²⁷⁾ En él se debe interpretar la formación de caminos, acompañado por los índices de implicación en el sentido del análisis implicativo, clásico o entrópico, según la opción elegida. Estas autoras consideran que el modo cono de este gráfico resulta el más informativo para el objetivo que se persigue. Así en un estudio de factores de riesgo, el modo cono permite identificar los factores de riesgo, cuando en el cono del gráfico se coloca la variable que representa la enfermedad, y los factores protectores, cuando se sitúa la variable de no enfermedad, mientras que en un estudio de factores pronósticos se identifican los factores de buen pronóstico, cuando en el cono del gráfico se ubica el evento final favorable, por ejemplo, vivo y los factores de mal pronóstico, cuando se ubica el evento final desfavorable, por ejemplo, fallecido.

Los métodos gráficos tienen ventaja sobre los métodos de regresión, ya que estos últimos a veces

excluyen confusores importantes o dejan en el modelo "optimizado" variables erróneamente clasificadas como confusores, dado que la selección se realiza a través de algoritmos matemático-estadísticos, basados en el valor p, donde la influencia del investigador es mínima; en cambio los DAG son construidos sobre el conocimiento experto a priori del investigador, minimizando la introducción de sesgos durante el diseño del estudio y en el análisis de resultados. ⁽¹⁰⁾

Los DAG permiten identificar los factores que influyen sobre un desenlace, pero no permiten identificar si influyen a favor o en contra de este. Sin embargo, autoras de este trabajo consideran que mediante el grafo implicativo en modo cono, del análisis estadístico implicativo, se puede reconocer cuando un factor es protector o de riesgo. Los DAG tampoco dan información sobre la magnitud del efecto, mientras un grafo implicativo indica el nivel de intensidad de todas las implicaciones. Estas ventajas se evidencian en los trabajos recientemente realizados por estas autoras, con el empleo del análisis estadístico implicativo en investigaciones médicas. ⁽²⁷⁻³¹⁾

Cuando se trata con variables modales, a este gráfico se le denomina grafo de propensión. ⁽³²⁾ Otro gráfico propio del análisis estadístico implicativo es el árbol cohesitivo.

El árbol cohesitivo

El árbol cohesitivo, o de cohesión, es la representación gráfica de las reglas generadas en el análisis estadístico implicativo. La cohesión permite estructurar el conocimiento en forma de reglas y meta-reglas. Al ser de carácter no lineal, asimétrico, jerárquico y dinámico permite superar la simple articulación de las partes de una tipología clásica, a fin de alcanzar un todo significativo. ⁽³³⁻³⁵⁾

Las reglas y meta-reglas que surgen en este análisis se pueden presentar en tres esquemas:

$R \Rightarrow c$, donde $R : a \Rightarrow b$, que se interpreta como que c es consecuencia de la regla R;

$a \Rightarrow R$, donde $R : b \Rightarrow c$, que se interpreta como que a se dedujo de la regla R o que la regla R es consecuencia de a;

$R_1 \Rightarrow R_2$, donde $R_1 : a \Rightarrow b$ y $R_2 : c \Rightarrow d$, que se interpreta como que R_2 se dedujo de la regla R_1 o que la regla R_2 es consecuencia de R_1 .

Intuitivamente, la cohesión mide el desequilibrio de las frecuencias de los eventos $\neg a \vee b$ y $a \wedge \neg b$ a favor del primero.

En este gráfico las reglas y meta-reglas se estructuran de manera ascendente en clases ajustadas y orientadas. En cada nivel se calcula el índice de cohesión entre cada par (ordenado) de clases del nivel anterior y se forma una nueva clase que reúne (y reemplaza) a las dos anteriores y así sucesivamente. El índice de cohesión mide la fuerza de la consistencia de las variables involucradas en la

clase creada. Las parejas de variables, que forman parte de una misma clase, tienen índices de cohesión, generalmente, mayores a los de las parejas de variables que todavía se mantienen en clases distintas, lo que se evidencia en los trabajos recientes en las ciencias médicas de estas autoras. ⁽²⁷⁻³¹⁾

Gráficos que apoyan los métodos de análisis multivariados

Una etapa importante en el análisis y modelación de datos multivariados es su representación gráfica. Los métodos gráficos apoyan visualmente a las técnicas multivariadas y forman parte indispensable del análisis. A continuación, se describen algunos ejemplos.

En el análisis factorial exploratorio y el análisis de componentes se emplean:

- El gráfico de sedimentación de las componentes, donde se observa la relación entre los autovalores y el número de componentes. El mismo suele ser utilizado, también, como contraste gráfico para conocer el número de componentes o factores a retener en un análisis factorial exploratorio o un análisis de componentes principales. Según este criterio, se retienen todas las componentes que están situadas previamente a la parte del gráfico en la que los componentes empiezan a no presentar pendientes fuertes. ⁽³⁶⁾

- El biplot, que muestra las observaciones y las variables en el mismo gráfico, de forma tal que se puedan hacer interpretaciones sobre las relaciones conjuntas. Las observaciones son generalmente graficadas como puntos. La configuración de los puntos es obtenida a partir de combinaciones lineales de las variables originales. Las variables son graficadas como vectores desde el origen. Los ángulos entre las variables representan la correlación entre las variables. Sirve para representar las componentes encontradas en un análisis de componentes principales. ⁽³⁷⁾

- La gráfica de influencias empleada para identificar cuáles variables tienen el mayor efecto en los factores. Se plotan en un plano cartesiano diferentes rotaciones, de manera que se pueda interpretar con más facilidad las influencias de los factores. Las influencias que se aproximan a -1 o 1 indican que la variable afecta considerablemente al factor. Las influencias cercanas a 0 indican que la variable tiene poca influencia en el factor. ⁽³⁶⁾

- La gráfica de doble proyección empleada para evaluar la estructura de los datos y las influencias de los dos primeros factores. Se relacionan en un plano cartesiano las puntuaciones del segundo factor, versus las puntuaciones del primer factor. ⁽³⁶⁾

Otros diagramas empleados como apoyo a otras técnicas multivariadas son:

- El triplot, variante del gráfico biplot, sobre el que además se grafican covariables para explicar la asociación entre los marcadores filas y columnas de la matriz de variables dependientes, representados en el biplot. Así las tres dimensiones del análisis, es decir, los casos, las variables respuestas y las

variables predictoras, medidas sobre los mismos casos, serán representadas en un mismo punto. Se emplea en el análisis de correlaciones canónicas, ya que permite ilustrar la correlación entre matrices. ⁽³⁷⁾

- El diagrama de escalamiento multidimensional, útil para mostrar distancias entre datos para los cuales las medidas euclídeas no son apropiadas. El objetivo es mostrar las relaciones entre observaciones, representadas por distancias o similitudes, en un plano tal que las distancias verdaderas sean preservadas tanto como sea posible. ⁽³⁷⁾

- El árbol de recorrido mínimo, el cual se construye uniendo puntos que representan observaciones multivariadas y que se proyectan en un plano como resultado de alguna técnica de reducción de dimensión. Los puntos son conectados con segmentos de líneas rectas tales, que todos los puntos quedan unidos directa o indirectamente, de manera que la suma de las longitudes de todos los segmentos es mínima. ⁽³⁷⁾

- El diagrama de secuencia, que se construye previo a la formación de las ecuaciones del modelo de ecuaciones estructurales, una de las técnicas más certeras en el estudio de la causalidad. En él se grafican las relaciones, quedando las variables observadas representadas por cuadrados y las latentes por círculos. ⁽³⁸⁾

- El diagrama de influencia, o de relevancia, que se hace previo a los árboles de decisión para visualizar las relaciones entre variables o el modelo de un sistema y ganar en claridad en la toma de decisiones complejas. En él las influencias externas se enlazan con las variables de decisión o las intermedias.

- El dendograma que muestra las agrupaciones que se forman a partir de un análisis de conglomerados.

- El árbol de similaridad, que acompaña las técnicas del análisis estadístico implicativo, es similar al dendograma y se construye con el mismo objetivo, pero, a diferencia de este, en el cual todas las variables o individuos quedan agrupados, solo se representan las agrupaciones significativas y se obtiene en base al índice de similaridad de Lerman, que no es simétrico.

- Los árboles de regresión que son un conjunto de reglas determinadas por un procedimiento de ajuste por particiones binarias recursivas, donde un conjunto de datos es sucesivamente particionado y sirven para representar cualquier problema, que puede ser canalizado con regresión lineal donde la variable dependiente es numérica. ⁽³⁷⁾

- Los árboles de clasificación, que funcionan como los árboles de regresión, pero en este caso la variable dependiente es categórica, por lo que sirven para representar cualquier problema que puede ser canalizado con el análisis discriminante o la regresión logística. ⁽³⁷⁾

- Los gráficos de control multivariado, como el Shewhart multivariado o T2 de Hotelling, el MEWMA (gráfico multivariado de promedios móviles ponderados exponencialmente) y el MCUSUM (gráfico multivariado de sumas acumuladas).

- Los gráficos para la validación de los modelos obtenidos, como la curva de características operativas del receptor, conocida por su sigla en

inglés ROC y las curvas de calibración bivariadas y multivariadas.

El investigador desde su creatividad puede variar algunos aspectos de las gráficas y añadir nuevos elementos que favorezcan el entendimiento. Por ejemplo, Persano ⁽³⁹⁾ y colaboradores emplean gráficos de similaridad, donde plotean las unidades de análisis que forman diferentes grupos, en diferentes momentos, pudiendo "exportar gráficamente" los datos de un diagrama de barras de error, apiladas en cuyo eje de las "Y" se encuentra el índice de similaridad promedio hacia este tipo de gráfico.

Sin dudas, el apoyo visual de estos gráficos es indispensable para tomar decisiones en el análisis por cuanto a simple vista se arriban a conclusiones de una manera ágil, que resultarían complejas por la cantidad de datos a procesar.

Previo a este análisis multivariado explicativo, es preciso explorar los datos disponibles. Igual que existen técnicas descriptivas univariadas y bivariadas, existen las multivariadas que sirven para describir el comportamiento de un conjunto de datos, donde puede haber múltiples variables tanto dependientes como independientes. Uno de los métodos más empleados en el análisis exploratorio de datos son los gráficos.

Gráficos para la exploración de datos

Los gráficos, además de servir para la presentación de datos y resultados, permiten identificar observaciones raras, tendencias y/o agrupamientos preliminares e hipotetizar posibles modelos para su análisis. Debido al gran número de variables involucradas, no es una tarea fácil examinar gráficamente datos multivariados. Además de realizar histogramas, gráficos de cajas (*box-plot*) y de tallos y hojas, sobre las observaciones (preferentemente estandarizadas) de cada variable, existen diversos tipos de gráficas para representar observaciones multivariadas, muchas de ellas son verdaderas técnicas de reducción de dimensión. En general, se persigue la graficación en espacios visuales de menor dimensión que preservan las relaciones entre las observaciones en el espacio original. ⁽³⁷⁾

Algunos de los diagramas empleados en relaciones bivariadas también pueden ser empleados en el análisis multivariado, por ejemplo, el diagrama de dispersión permite la representación de las dos primeras componentes en un análisis de componentes principales, o de los dos primeros ejes canónicos obtenidos de un análisis discriminante lineal, o la relación entre las dos primeras variables canónicas de un análisis de correlaciones canónicas. ⁽³⁷⁾

Estos diagramas bidimensionales pueden tornarse tridimensionales, como los diagramas de dispersión, cuando se plotean en un espacio tridimensional las tripletas de valores correspondientes a tres variables, o el diagrama de burbujas en un plano cartesiano, donde cada eje corresponde a una

variable y la tercera variable queda representada por el tamaño de la burbuja.

Existen otros diagramas propiamente multidimensionales como:

- La matriz de gráficos de dispersión: matriz de gráficos de dispersión de todos los pares de variables en estudio, en cuya diagonal principal pueden aparecer histogramas, gráficos de cajas o simplemente el nombre de cada variable según su orden. También, se pueden introducir variables de agrupamiento a través de la selección de colores y formas de símbolos en cada gráfico de dispersión. ⁽³⁷⁾

- Las caras de Chernoff: representación del conjunto de variables mediante caras, asociando el comportamiento de las variables que intervienen en el estudio con características faciales diferentes, como la posición, longitud, altura, inclinación y forma de ojos, cejas, nariz, boca, orejas, pelo y otros componentes faciales, etc., por ejemplo, la abertura de la boca, la sonrisa. ^(39,40) Una variante de este son las caras asimétricas que proponen una nueva cara, en la cual los parámetros del lado derecho de la cara pueden variar independientemente de los parámetros del lado izquierdo, con lo que se pueden representar hasta 36 variables, en lugar de las 18 variables originales de Chernoff.

- El diagrama de estrellas: aplicable, si todas las variables toman valores positivos y consiste en graficar rayos o ejes que parten de un punto central para cada variable y donde su longitud corresponde al valor de la variable. ⁽⁶⁾ Para eliminar valores negativos, se puede seleccionar como centro del círculo al valor más pequeño y expresar los otros valores en relación a este. Otra técnica usada con el mismo fin es el uso de la transformación de rango. Las estrellas son usualmente presentadas en arreglo matricial. ⁽⁴⁰⁾

- El diagrama de radar: semejante a un gráfico de estrella, pero en él cada estrella representa cada una de las variables en estudio y no una unidad de observación; las unidades de observación identifican rayos concéntricos. Los valores individuales de cada variable, debidamente estandarizados se representan sobre los rayos y la conexión de ellos genera la estrella de esa variable. Este es un tipo de gráfico de estrellas formado por un conjunto de rayos a partir de un centro, determinados por una variable categórica seleccionada previamente y un conjunto de estrellas concéntricas, definidas por el comportamiento de cada una de variables numéricas en un espacio específico de datos. El espacio determinado es establecido por la unidad de referencia de la variable en todos los subgrupos. La longitud de un determinado rayo es proporcional a la magnitud de la variable para el subgrupo con respecto al máximo de la variable a través de todos los subconjuntos. ⁽⁴⁰⁾

- Las curvas de Andrews: método basado en la transformación de Fourier para representar datos multivariados en dos dimensiones, donde cada punto multidimensional es representado por una función y se construye graficando todas las funciones que representan a la observación *i*-ésima de un vector aleatorio. Cada variable de cada observación se

representa por una componente individual en la suma de la transformada de Fourier. La magnitud de cada variable de un sujeto o elemento particular afecta la frecuencia, la amplitud y la periodicidad de la función, dando una representación única para cada sujeto. ⁽⁴⁰⁾

- El diagrama de perfiles multivariados: permiten visualizar los perfiles multivariados de cada sujeto cuando se tienen medidas de una variable, repetidas en el tiempo, o varias variables, medidas en un mismo sujeto o unidad experimental. Si los perfiles que se muestran son el resultado de promediar la respuesta de varias repeticiones o unidades, entonces tiene sentido mostrar barras de error estándar para cada variable. ⁽³³⁾

Todos estos gráficos permiten detectar valores atípicos o extremos y requieren la estandarización de las variables a representar, o sea, que estén medidas en iguales unidades y habrá que tener en cuenta, además, el orden de las variables que puede afectar la interpretación.

Gráficos de análisis espacial

Otro aspecto a estudiar en la multicausalidad es la dimensión espacial. Entre los métodos gráficos para su estudio se destacan: ⁽³⁷⁾

- El semivariograma formado a partir de una variable sintética resumen.

- Los mapas sintéticos creados mediante la interpolación de los valores, obtenidos por la predicción de los valores de una variable aleatoria sobre un conjunto de puntos definidos por una localización en el espacio.

- Las redes de conexión o gráficos de vecindario que se generan conectando individuos vecinos en un mapa mediante la triangulación de Delaunay, cuando las entidades se encuentran distribuidas en forma homogénea en el espacio, incluyendo entidades periféricas que no deberían estar relacionadas.

- El gráfico de Gabriel, subconjunto del gráfico de Delaunay que no incluye las conexiones periféricas.

Estas redes de conexión pueden ser adaptadas, pudiéndose excluir contactos entre puntos cercanos o incluir relaciones entre puntos lejanos, siguiendo criterios biológicos.

CONCLUSIONES

Los métodos gráficos son de gran utilidad, ya que posibilitan la representación de las relaciones entre las múltiples variables del estudio y, por otro lado, la exploración del conjunto de datos para identificar posibles valores extremos, agrupaciones por semejanzas, etc. En especial, los diagramas causales permiten visualizar las relaciones multivariadas, detectando inclusive los factores confusores. Dentro de estos se destaca el grafo implicativo, que, al igual que un grafo acíclico orientado, permite identificar los factores que influyen sobre un desenlace, pero superándolo en dos sentidos, ya que puede distinguir cuando se trata de un factor de riesgo o protector y, por otro lado, permite conocer cuál es la magnitud del efecto, a través de la intensidad implicativa.

REFERENCIAS BIBLIOGRÁFICAS:


1. Velarde Martínez A. Generación de Grafos Dirigidos Acíclicos mediante Métodos Aleatorios para la Planificación de Tareas. *Conciencia Tecnológica* [revista en internet]. 2018 [citado 3 de abril 2019]; 2018(55): 5-13. Disponible en: <http://www.redalyc.org/jatsRepo/944/94455712004/index.html>.
2. Textor J, van der Zander B, Gilthorpe MS, Liškiewicz M, Ellison G TH. Robust causal inference using directed acyclic graphs: the R package 'dagitty'. *International Journal of Epidemiology* [revista en internet]. 2016 [citado 3 de abril 2019]; 45(6): 1887-1894. Disponible en: <https://doi.org/10.1093/ije/dyw341>.
3. Arnold KF, Harrison WJ, Heppenstall AJ, Gilthorpe MS. DAG-informed regression modelling, agent-based modelling and microsimulation modelling: a critical comparison of methods for causal inference. *International Journal of Epidemiology* [revista en internet]. 2019 [citado 3 de abril 2019]; 48(1): 243-253. Disponible en: <https://academic.oup.com/ije/advance-article/doi/10.1093/ije/dyy260/5231935>.
4. Rodrigues Cortes T, Faerstein E, Struchiner CJ. Utilização de diagramas causais em epidemiologia: um exemplo de aplicação em situação de confusão. *Cad. Saúde Pública* [revista en internet]. 2016 [citado 3 de abril 2019]; 32(8): e0013115. Disponible en: <http://dx.doi.org/10.1590/0102-311X0013115>.
5. Brewer LE, Wright JM, Rice G, Neas L, Teuschler L. Causal inference in cumulative risk assessment: The roles of directed acyclic graphs. *Environment International* [revista en internet]. 2017 [citado 3 de abril 2019]; 102: 30-41. Disponible en: <http://dx.doi.org/10.1016/j.envint.2016.12.005>.
6. Greenland S, Pearl J. Causal diagrams. *Wiley Stats Ref* [en línea]. EUA: Statistics Reference Online, John Wiley & Sons [citado 3 de abril 2019]. Disponible en: <https://doi.org/10.1002/9781118445112.stat3732.pub2>.
7. Williamson EJ, Aitken Z, Lawrie J, Dharmage SC, Burgess JA, Forbes AB. Introduction to causal diagrams for confounder selection. *Respirology* [revista en internet]. 2014 [citado 3 de abril 2019]; 19(3): 33-311. Disponible en: <https://onlinelibrary.wiley.com/doi/full/10.1111/resp.12238>.
8. Hernan M, Robins J. Causal inference [en línea]. Boca Raton: Chapman & Hall/ CRC, forthcoming; 2019. [citado 3 de abril 2019]. Disponible en: <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>.
9. Pearl J. Comments on: The tale wagged by the DAG. *International Journal of Epidemiology* [revista en internet]. 2018 [citado 3 de abril 2019]; 47(3): 1002-1004. Disponible en: https://ftp.cs.ucla.edu/pub/stat_ser/r474.pdf.
10. Werlinger F, Cáceres DD. Aplicación de grafos acíclicos dirigidos en la evaluación de un set mínimo de ajuste de confusores: un complemento al modelamiento estadístico en estudios epidemiológicos observacionales. *Rev. méd. Chile* [revista en internet]. 2018 [citado 3 de abril 2019]; 146(7): 907-913. Disponible en: <http://dx.doi.org/10.4067/s034-98872018000700907>.
11. Ananth CV, Schisterman EF. Confounding, causality and confusión: the role of intermediate variables in interpreting observational studies in obstetric. *Am J Obstet Gynecol* [revista en internet]. 2017 [citado 3 de abril 2019]; 217(2): 167-175. Disponible en: <http://dx.doi.org/10.1016/j.ajog.2017.04.016>.
12. Mansournia MA, Higgins JPT, Sterne JAC, Hernán MA. Biases in Randomized Trials A Conversation Between Trialists and Epidemiologists. *Epidemiology* [revista en internet]. 2017 [citado 3 de abril 2019]; 28(1): 54-9. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5130591/>.
13. Werneck GL. Diagramas causales: la epidemiología brasileña regresa al futuro. *Cad. Saúde Pública* [revista en internet]. 2016 [citado 3 de abril 2019]; 32(8): e00120416. Disponible en: <http://dx.doi.org/10.1590/0102-311X00120416>.
14. Williams TC, Bach CC, Matthiesen NB, Henriksen TB, Gagliardi L. Directed acyclic graphs: a tool for causal studies in paediatrics. *Pediatric Research* [revista en internet]. 2018 [citado 3 de abril 2019]; 84: 487-493. Disponible en: <https://doi.org/10.138/s41390-018-0071-3>.
15. Pearce N, Lawlor DA. Causal inference—so much more than statistics. *International Journal of Epidemiology* [revista en internet]. 2016 [citado 3 de abril 2019]; 45(6): 1895-193. Disponible en: <https://doi.org/10.1093/ije/dyw328>.
16. Lousdal ML. An introduction to instrumental variable assumptions, validation and estimation. *Emerging Themes in Epidemiology* [revista en internet]. 2018 [citado 3 de abril 2019]; 15: 1. Disponible en: <https://doi.org/10.1186/s12982-018-0069-7>.
17. Moreno Betancur M, Lee KJ, Leacy FP, White IR, Simpson JA, Carlin JB. Canonical Causal Diagrams to Guide the Treatment of Missing Data in Epidemiologic Studies. *American Journal of Epidemiology* [revista en internet]. 2018 [citado 3 de abril 2019]; 187(12): 2705-2715. Disponible en: <https://academic.oup.com/aje/article-abstract/187/12/2705/5074351>.
18. Tirado Ríos NR, Triana Litardo FE, Saa Saltos JW. Optimización de Redes Bayesianas basado en técnicas de aprendizaje por inducción. *Revista Publicando* [revista en internet]. 2016 [citado 3 de abril 2019]; 3(9): 41-60. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=5833529>.

19. Monsalve Torra AE. Sistema de ayuda a la decisión clínica en enfermedades de diagnóstico complejo [tesis]. España: Universidad de Alicante; 2017 [citado 3 de abril 2019]. Disponible en: https://rua.ua.es/dspace/bitstream/10045/65334/1/tesis_monsalve_torra.pdf.
20. D´Negri CE, De Vito EL. Introducción al razonamiento aproximado: lógica difusa. *Revista Argentina de Medicina Respiratoria* [revista en internet]. 2006 [citado 3 de abril 2019]; 6(3): 126-136. Disponible en: <http://www.redalyc.org/pdf/3821/382138367007.pdf>.
21. Berlanga Silvente V, Rubio Hurtado MJ, Vilà Baños R. Cómo aplicar árboles de decisión en SPSS. *REIRE* [revista en internet]. 2013 [citado 3 de abril 2019]; 6(1): 65-79. Disponible en: <http://diposit.ub.edu/dspace/bitstream/2445/43762/1/618361.pdf>.
22. Leyva Vázquez M, Pérez Teruel K, Febles Estrada A, Gulín González J. Técnicas para la representación del conocimiento causal: un estudio de caso en Informática Médica. *Revista Cubana de información en ciencias de la salud* [revista en internet]. 2013 [citado 3 de abril 2019]; 24(1): 73-83. Disponible en: <http://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=40535>.
23. Mar Cornelio O, Leyva Vázquez M, Santana Ching I. Modelo multicriterio multiexperto utilizando Mapa Cognitivo Difuso para la evaluación de competencias. *Ciencias de la Información* [revista en internet]. 2015 [citado 13 Ene 2019]; 46(2): 17-22. Disponible en: <http://www.redalyc.org/pdf/1814/181441052004.pdf>.
24. Valls Pla X, Gregori Huerta P. Diseño de un paquete R para el Análisis Estadístico Implicativo Trabajo Final de Master en Matemática Computacional [tesis]. España: Universitat Jaume I; 2014. [citado 10 de enero 2019]. Disponible en: <http://repositori.uji.es/xmlui/handle/10234/107441>.
25. Couturier R, Pazmiño R. Use of Statistical Implicative Analysis in Complement of Item Analysis. *International Journal of Information and Education Technology* [revista en internet]. 2016 [citado 22 de enero 2019]; 6(1): 39-43. Disponible en: <http://www.ijiet.org/vol6/655-DL0026.pdf>.
26. Marín Martínez P, Gregori Huerta P. Probability distribution of the classical implication intensity seen as a random variable in Statistical Implicative Analysis. Master's Degree in Computational Mathematics [tesis]. España: Universitat Jaume I de Castellón; 2017 [citado 22 de enero 2019]. Disponible en: http://repositori.uji.es/xmlui/bitstream/handle/10234/171834/TFM_2017_MarinMartinez_Paloma.pdf?sequence=1.
27. García Mederos Y, Zamora Matamoros L, Sagaró del Campo N. Análisis estadístico implicativo en la identificación de factores de riesgo en pacientes con cáncer de pulmón. *MEDISAN*. [revista en internet]. 2015 [citado 22 de enero 2019]; 19(8): 947-957. Disponible en: <http://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=60754>.
28. Moraga Rodríguez A, Zamora Matamoros L, Sagaró del Campo NM, Moraga Rodríguez A, Rodríguez Griñán A. Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de pulmón. *MEDISAN* [revista en internet]. 2016 [citado 22 de enero 2019]; 20(3). Disponible en: <http://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=64671>.
29. Moraga Rodríguez A, Zamora Matamoros L, Sagaró del Campo NM, Moraga Rodríguez A, Rodríguez Griñán A. Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de mama. *MEDISAN* [revista en internet]. 2017 [citado 22 de enero 2019]; 21(4): 395-406. Disponible en: <http://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=71774>.
30. Moraga Rodríguez A, Zamora Matamoros L, Sagaró del Campo NM, Moraga Rodríguez A, Rodríguez Griñán A. Análisis estadístico implicativo para la identificación de factores pronósticos de la mortalidad por cáncer de próstata. *MEDISAN* [revista en internet]. 2018 [citado 22 de enero 2019]; 22(1). Disponible en: <http://www.medigraphic.com/cgi-bin/new/resumen.cgi?IDARTICULO=78550>.
31. Paez Candelaria Y, Sagaró del Campo NM, Zamora Matamoros L. Análisis estadístico implicativo en la determinación de factores pronósticos del estado nutricional del paciente grave al egreso. *MEDISAN* [revista en internet]. 2018 [citado 22 de enero 2019]; 22(6). Disponible en: <http://www.medisana.sld.cu/index.php/san/article/view/2002>.
32. Zamora Matamoros L, Díaz Silvera J. Estudio de relaciones causales entre indicadores de promoción del primer año de la carrera Ciencia de la Computación con el análisis implicativo modal. *Maestro y Sociedad* [revista en internet]. 2018 [citado 22 de enero 2019]; 15(2): 203-212. Disponible en: <https://revistas.uo.edu.cu/index.php/MyS/article/view/3520>.
33. Gras R. Genese et developpement de l'analyse statistique implicative: retrospective Historique. *Educ. Matem. Pesq.* [revista en internet]. 2014 [citado 22 de enero 2019]; 16(3): 645-661. Disponible en: https://revistas.pucsp.br/emp/article/viewFile/21541/pdf_1.
34. Gras R, Régnier JC, Lahanier-Reuter D, Marinica C, Guillet F. L'Analyse Statistique Implicative. *Des Sciences dures aux Sciences Humaines et Sociales 3ème* [en línea] Cépaduès: revue et augmentée; 2017 [citado 10 de enero 2019]. Disponible en: <http://critiquelittera.info/lanalyse-statistique-implicative-des-sciences-dures-aux-sciences-humaines-et-sociales-pdf-et-epub-rgis-gras.pdf>.

35. Couturier R, Gregori P. Classification en utilisant les règles d'implication de l'ASI. ResearchGate [revista en internet]. 2017 [citado 10 de enero 2019]. Disponible en: <https://www.researchgate.net/publication/324770682>.
36. MINITAB. Interpretar todos los estadísticos y gráficas para análisis factorial [en línea]. México: Soporte Técnico de Minitab 18; 2019 [citado 10 de enero 2019]. Disponible en: <https://support.minitab.com/es-mx/minitab/18/help-and-how-to/modeling-statistics/multivariate/how-to/factor-analysis/interpret-the-results/all-statistics-and-graphs/>.
37. Balzarini M, Bruno C, Córdoba M, Teich I. Herramientas en el Análisis Estadístico Multivariado. Escuela Virtual Internacional CAVILA Facultad de Ciencias Agropecuarias. Argentina: Universidad Nacional de Córdoba; 2015. Disponible en: https://www.researchgate.net/profile/Monica_Balzarini3/publication/286931204_Herramientas_en_el_Analisis_Estadistico_Multivariado/links/5670e57808ae0d8b0cc1001a/Herramientas-en-el-Analisis-Estadistico-Multivariado.pdf.
38. Keith TZ. Multiple regression and beyond: An introduction to multiple regression and structural equation modeling. 2nd ed. New York: Taylor & Francis; 2015. Disponible en: <https://www.taylorfrancis.com/books/9781317608066>.
39. Persano Adorno D, Pizzolato N, Fazio C. Long term stability of learning outcomes in undergraduates after an open-inquiry instruction on thermal science. Physical Review Physics Education Research [revista en internet]. 2018 [citado 22 de enero 2019]; 14(1) Disponible en: <https://journals.aps.org/prper/abstract/10.1103/PhysRevPhysEducRes.14.010108>.
40. Jacoby W. Statistical graphic for visualizing multivariate data. Sage Publication Inc. EUA: Sage University paper; 1998 [citado 15 de enero 2019]. Disponible en: <http://methods.sagepub.com/book/statistical-graphics-for-visualizing-multivariate-data>.
41. Schiattino I, Silva C. Representación gráfica de información multivariante. Aplicación al sistema de salud de Chile. Rev Chil Salud Pública [revista en internet]. 2013 [citado 20 de enero 2019]; 17(2): 117-123. Disponible en: <https://revistasaludpublica.uchile.cl/index.php/RCSP/article/download/27126/28758/>.


Declaración de autoría

Nelsa María Sagaró-del-Campo

 <https://orcid.org/0000-0002-1964-8830>

Participó en la concepción y diseño del artículo, en la revisión crítica de la literatura y la elaboración de las versiones original y final.

Larisa Zamora-Matamoras

 <https://orcid.org/0000-0003-2210-0806>

Participó en la concepción y diseño del artículo, en la revisión crítica de la literatura y la elaboración de las versiones original y final.

Los autores declaran que no existen conflictos de intereses.

Copyright Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta. Este artículo está bajo una [licencia de Creative Commons Reconocimiento 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/), los lectores pueden realizar copias y distribución de los contenidos por cualquier medio, siempre que se mantenga el reconocimiento de sus autores.