



ARTÍCULO DE REVISIÓN

Técnicas estadísticas multivariadas para el estudio de la causalidad en Medicina

Multivariate statistical techniques for the study of causality in medicine

Nelsa María Sagaró del Campo¹, Larisa Zamora Matamoros²

¹Universidad de Ciencias Médicas. Santiago de Cuba, Cuba.

²Universidad de Oriente. Santiago de Cuba, Cuba.

Recibido: 17 de abril 2019

Aceptado: 12 de diciembre 2019

Publicado: 15 de abril de 2020

Citar como: Sagaró del Campo NM, Zamora Matamoros L. Técnicas estadísticas multivariadas para el estudio de la causalidad en Medicina. Rev Ciencias Médicas [Internet]. 2020 [citado: fecha de acceso]; 24(2): e4029. Disponible en: <http://revcmpinar.sld.cu/index.php/publicaciones/article/view/4029>

RESUMEN

Introducción: las técnicas estadísticas multivariadas son aquellas que analizan múltiples características medidas en un mismo individuo, que por estar interrelacionadas no tiene sentido medir su efecto de manera aislada. Las mismas son más frecuentemente empleadas en los estudios observacionales acerca de la etiología y el pronóstico de una enfermedad, ya que permiten explicar, predecir, controlar variables confusoras y obtener mayor consistencia en la inferencia estadística.

Objetivo: exponer las técnicas multivariadas que pueden ser empleadas para el estudio de la causalidad en las ciencias biomédicas.

Métodos: se llevó a cabo una revisión de la literatura sobre el tema en bases de datos bibliográficas como Pubmed/Medline, SciELO, SCOPUS, Web of Science, EBSCOhost, Google, entre otras. Se emplearon en la estrategia de búsqueda como descriptores los consignados en las palabras claves del artículo, los mismos fueron utilizados en idiomas inglés, francés, portugués y español, combinados con los operadores lógicos. De los más de 200 documentos digitales encontrados, se escogieron los de mayor actualidad, elaborando un informe donde se agrupan las técnicas según su objetivo y tipos de variables empleadas en el análisis.

Resultados: se presentan las técnicas descriptivas, explicativas y mixtas, considerando los diferentes tipos de regresión y las ecuaciones estructurales como las técnicas más empleadas en el estudio de la causalidad.

Conclusiones: se propone el empleo del análisis estadístico implicative en la determinación de factores de riesgo y pronósticos.

Palabras clave: Estadísticas; Análisis Multivariante; Análisis De Regresión; Análisis Estadístico.

ABSTRACT

Introduction: multivariate statistical techniques are those that examine multiple characteristics considered in the same individual, and that because they are interrelated do

not make sense to determine their effect in an isolated way. They are more frequently applied in observational studies in relation to the etiology and prognosis of a disease, since they permit to explain, predict and control confusing variables to obtain greater consistency in statistical inference.

Objective: this paper is aimed at describing the multivariate techniques that can be applied for the study of causality in biomedical sciences.

Methods: a review of the literature on the subject was carried out, searching in bibliographic databases such as Pubmed/Medline, SciELO, SCOPUS, Web of Science, EBSCOhost, Google, among others. The descriptors used in the search strategy were those provided in the key words of the article, which were used in English, French, Portuguese and Spanish, combined with the logical operators. Out of more than 200 digital documents found, the most current ones were chosen, and a report was prepared where the techniques were grouped according to their objective and types of variables applied in the analysis.

Results: descriptive, explanatory and mixed techniques were presented, considering the different types of regression and structural equations as the most applied techniques in the study of causality.

Conclusions: The use of implicit statistical analysis in the determination of risk factors and prognosis was proposed.

Keywords: Statistics; Multivariate Analysis; Regression Analysis; Statistical Analysis.

INTRODUCCIÓN

Las técnicas estadísticas multivariadas son aquellas que analizan múltiples características, medidas en un mismo individuo. Constituyen una generalización de las técnicas univariadas y bivariadas, donde todas las variables deben ser aleatorias y estar interrelacionadas de tal forma que no tenga sentido interpretar de forma aislada sus diferentes efectos.⁽¹⁾

Los modelos de análisis multivariado son empleados con más frecuencia en los estudios observacionales acerca de la etiología y el pronóstico de una enfermedad, y en ocasiones en los estudios de diagnóstico o de intervención.⁽²⁾

El análisis multivariado está presente en los estudios de causalidad, cuya estrategia debe partir de un buen diseño, donde se involucren como variable dependiente el desenlace y como covariables los supuestos factores causales, donde son importantes los posibles factores confusores que deben incluirse en el análisis y se toman precauciones para evitar posibles sesgos.⁽³⁾

En los estudios observacionales, donde no es posible hacer una asignación aleatoria de los individuos a los diferentes grupos de estudio, es muy útil el análisis multivariado, ya que permite el ajuste de las características basales de los grupos para lograr una aproximación estadística razonable a la estimación del efecto de la intervención, una mayor economía en el análisis de los datos, hacer predicciones, el control de las variables confusoras y obtener una mayor consistencia en la inferencia estadística.

Las técnicas multivariadas son herramientas que permiten al investigador extraer abundante información de los datos disponibles. Las mismas son complejas y requieren para su utilización de un conocimiento profundo de sus fundamentos. Los paquetes estadísticos han permitido una amplia aplicación de las mismas, aunque algunas veces por insuficiente conocimiento y respaldo teórico, su empleo no ha sido el adecuado.⁽⁴⁾ Para el investigador novel que no está familiarizado con los términos empleados en el texto se sugiere consultar el anexo de terminología de uso frecuente.

La presente investigación tiene como objetivo exponer las técnicas multivariadas empleadas en el estudio de la causalidad en las ciencias biomédicas según su finalidad y tipo de variables.

MÉTODOS

Para lograr el objetivo propuesto se realizó una investigación documental con la revisión de las bases de datos biomédicas y la literatura apropiada por su carácter científico y actualizado. Se identificaron las regularidades y diferencias que permitieron la organización y elaboración de un material resumen de estas técnicas, donde se incluye el análisis estadístico implicative, empleado recientemente en las Ciencias Médicas.

se llevó a cabo una revisión de la literatura sobre el tema en bases de datos bibliográficas como Pubmed/Medline, SciELO, SCOPUS, Web of Science, EBSCOhost, Google, entre otras. Se emplearon en la estrategia de búsqueda como descriptores los consignados en las palabras claves del artículo, los mismos fueron utilizados en idiomas inglés, francés, portugués y español, combinados con los operadores lógicos. De los más de 200 documentos digitales encontrados, se escogieron los de mayor actualidad, elaborando un informe donde se agrupan las técnicas según su objetivo y tipos de variables empleadas en el análisis.

DESARROLLO

Para el estudio de la causalidad, al tener el paradigma multicausal que ha dominado desde el siglo anterior y la multicausalidad de los eventos en Medicina, las técnicas estadísticas apropiadas son las multivariadas, sobre todo las explicativas.

No obstante, en algunos casos se requiere de un primer análisis exploratorio, donde se hace necesario el empleo de las técnicas multivariadas descriptivas, para reducir dimensionalidad de los datos o agruparlos de manera conveniente; de ahí la utilidad de conocer todas las técnicas y en qué momento emplearlas.

Clasificación de las técnicas estadísticas multivariadas

Para saber qué técnica se debe aplicar, es necesario conocer cómo están medidas las variables que se analizan en el estudio.^(4,5) También es preciso conocer el tipo de relación existente entre las variables. La dependencia o no entre variables se determina por el conocimiento a priori del investigador. Como consecuencia de estos aspectos, los principales métodos de análisis multivariado empleados en la inferencia causal, se pueden agrupar según su finalidad (Fig. 1) en:

- **Métodos descriptivos o de interdependencia:** son empleados para verificar las relaciones de interdependencia y describir la estructura de los datos según se quieran identificar. Se dividen en:
 - variables similares
 - ✓ con variables cuantitativas se emplean el análisis de componentes principales y el análisis factorial exploratorio. Estas técnicas permiten reducir la dimensionalidad de los datos, con lo que se pueden descartar atributos irrelevantes para la toma de decisiones, que conspiran además en contra de la eficiencia de los algoritmos propios de las técnicas empleadas a posteriori.⁽⁵⁾ Las nuevas características encontradas son combinación lineal de las anteriores y se les llama componente si se forman por variables observadas y factores si se forman por variables no observadas.
 - ✓ con variables cualitativas se emplean el análisis de correspondencia y el modelo log-lineal para tablas de contingencia con más de dos entradas.

- individuos similares
 - ✓ con variables de cualquier tipo se emplean el análisis de conglomerado (*cluster*). Es una técnica de agrupamiento que clasifica las unidades de la matriz de datos en conglomerados (grupos internamente homogéneos y heterogéneos entre ellos) con el empleo de diferentes índices de similitud o de distancia. Estas clasificaciones pueden ser jerárquicas (ascendente o descendente) en los que no se establece a priori el número de grupos, o no jerárquicas en las que se parte de un número dado de grupos según hipótesis o cálculos previos.^(5,6)
- grupos de objetos similares
 - ✓ se emplea el escalamiento multidimensional, técnica que permite la reducción de datos, al tratar que la distancia entre los objetos en el nuevo espacio dimensional reducido sea lo más parecida a las distancias del espacio original, tal como la construcción de un mapa a menor escala.⁽⁵⁾
- **Métodos explicativos o de dependencia:** son empleados para analizar las relaciones de dependencia que permiten explicar o predecir determinados eventos. Según el tipo de variables dependientes pueden ser:
 - con variable dependiente cuantitativa
 - ✓ para explicar la variabilidad individual se puede emplear la regresión lineal múltiple, en el caso de una sola variable dependiente, o la correlación canónica y la regresión lineal multivariada en el caso de más de una variable dependiente.
 - ✓ para explicar la variabilidad grupal (los grupos condicionan la variable categórica) se emplea el análisis de varianza o de covarianza (ANOVA o ANCOVA) si es una sola variable o un análisis multivariado de varianza o de covarianza (MANOVA o MANCOVA) si se trata de más de una variable dependiente. En el ANCOVA y MANCOVA las variables explicativas pueden ser medidas en cualquier escala.⁽⁶⁾
 - con variable dependiente cualitativa:
 - ✓ el análisis discriminante para predecir si un individuo pertenece a un grupo cuando se cumple el supuesto de normalidad.
 - ✓ la regresión logística binaria para respuesta dicotómica o multinomial para respuesta politómica.⁽⁷⁾
 - ✓ el análisis conjunto (*conjoint*), que es el único entre los métodos multivariados en el cual el investigador construye primero un conjunto real o hipotético de variables, al combinar niveles escogidos de cada atributo.⁽¹⁾ Según el número de atributos se emplean variantes del método como los modelos auto-explicados y los modelos adaptativos e híbridos.
- **Métodos mixtos o estructurales** que combinan dependencia con interdependencia como las ecuaciones estructurales.

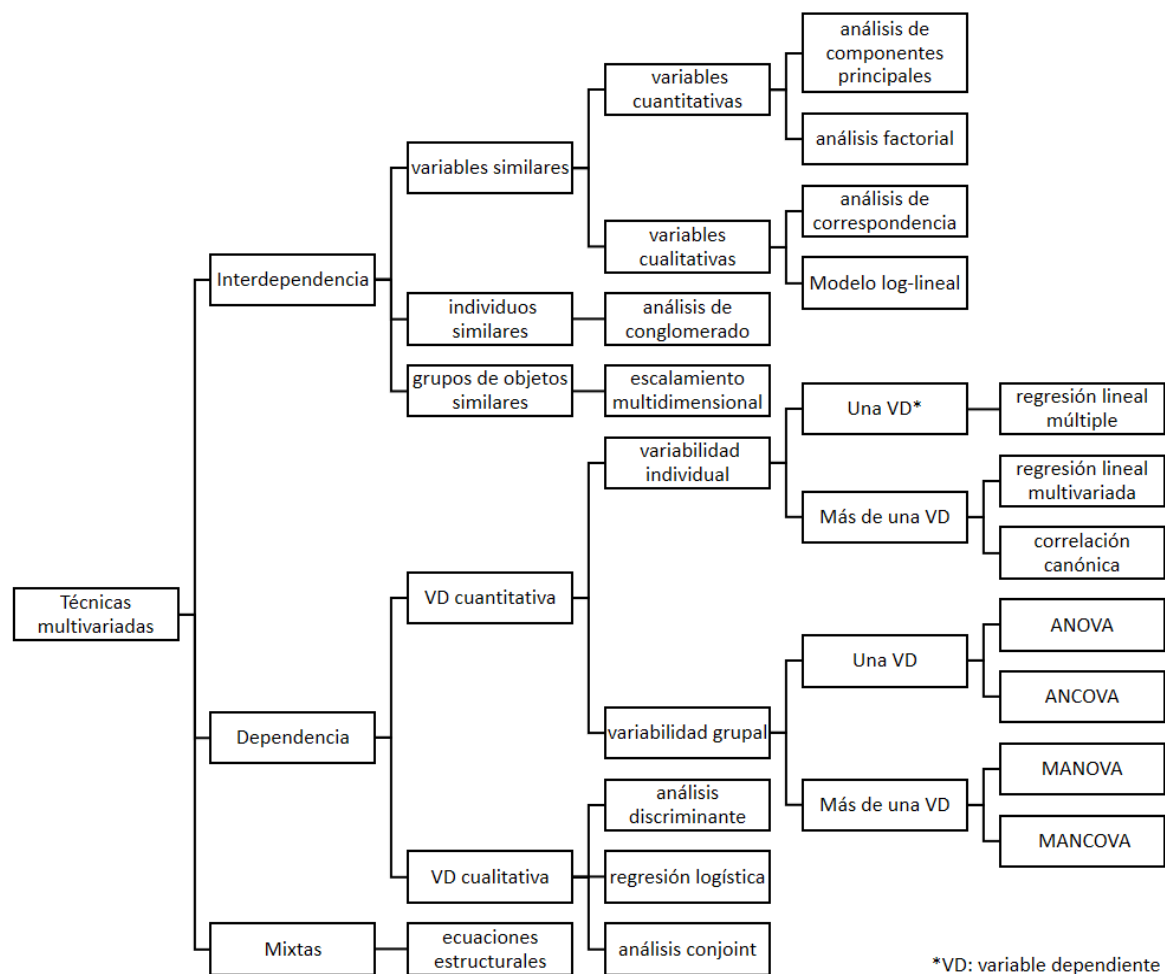


Fig. 1 Clasificación de las técnicas multivariadas

Las técnicas antes mencionadas pueden ser de amplio uso en la práctica, ya que en la clínica es frecuente emplear como variables, múltiples determinaciones de laboratorio como urea, creatinina, glucemia, colesterol, entre otras (variables cuantitativas) con la cual se pueden hacer investigaciones y análisis multivariados en dependencia de lo que se desea probar.

Por ejemplo, en un estudio donde se quiere evaluar la eficacia de cierto tratamiento para la Diabetes Mellitus se han medido un considerable número de variables que se corresponden con estas determinaciones del laboratorio clínico y se tiene interés en reducir su número. En este caso se puede emplear un análisis de componentes principales. (Método descriptivo o de interdependencia).

Si se quiere conocer si existen diferencia con respecto a estos parámetros del laboratorio entre varios grupos de individuos sometidos a diferentes dietas (factores que representan variables cualitativas), es posible emplear un MANOVA (Método explicativo o de dependencia).

Así, al analizar los tipos de variables y el objetivo que se quiere lograr, es posible identificar la técnica requerida en cada caso. No es tarea fácil decidir porque, en ocasiones, existe más de una técnica con la misma finalidad, por lo que es necesario conocer además qué pruebas son más robustas.

Al igual que existe evidencia de que las pruebas con una sola variable dependiente (regresión, ANOVA, ANCOVA, etc.) son robustas bajo la violación de los principios de normalidad y homocedasticidad (excepto cuando las muestras son muy pequeñas y desiguales); las técnicas multivariantes, con muestras grandes ($n > 30$) también son suficientemente robustas ante ligeras desviaciones de los supuestos estadísticos (excepto en algunas técnicas como los modelos de ecuaciones estructurales donde el tamaño muestral debe superior a las 100 unidades).⁽⁴⁾

Entre estos métodos multivariados, los más empleados para el estudio de la causalidad son la regresión múltiple y las ecuaciones estructurales.

Regresión múltiple

En los estudios observaciones, en los que no es posible asignar experimentalmente valores de una variable predictiva, al investigador le interesa conocer los efectos de múltiples variables predictivas sobre un resultado y para ello puede usar los modelos de regresión que sintetizan la información sobre variables predictivas individuales con el uso de todo el conjunto de datos. Los coeficientes estimados con el modelo son parámetros estadísticos descriptivos muy potentes que permiten obtener un significado de los datos.

Los métodos de regresión múltiple se emplean según el tipo de variable dependiente que se analiza, así se tienen, entre los más empleados:

- Regresión lineal: cuando la variable dependiente es continua. Los cambios de la covariable (aumento o disminución) influyen en el valor esperado de la variable dependiente de una manera lineal. Cada coeficiente de regresión representa el cambio que se espera en la variable de resultado por cada unidad de cambio en la covariable.⁽³⁾
- Regresión logística: cuando la variable dependiente es categórica. En su forma binaria es muy empleada en la identificación de factores de riesgo o pronósticos cuando la variable dependiente es dicotómica, por ejemplo, vivo o fallecido o enfermo o no enfermo, y se pretende encontrar la probabilidad de que ocurra el desenlace en estudio con la influencia o no de otras variables. También puede ser multinomial cuando permite tratar situaciones en las que la variable respuesta tienen más de dos de alternativas posibles, sean ordinales o no.⁽⁷⁾
- Regresión de Cox (Modelo de riesgos proporcionales): empleada en los estudios de supervivencia, donde la variable dependiente está en función del tiempo transcurrido hasta un determinado evento y se quiere determinar simultáneamente el efecto de una serie de factores que pueden alterar dicho tiempo. Ese efecto se presenta como la razón de riesgo (HR de *Hazard Ratio*) y expresa la magnitud en la que una variable aumenta o disminuye el riesgo de ocurrencia de un desenlace en el tiempo, por lo que supera a la técnica bivariada de Kaplan Meier.⁽⁸⁾
- Regresión binomial: empleada cuando la variable de respuesta es dicotómica, es decir, el resultado solo puede adoptar uno de dos valores posibles como enfermar o no enfermar, morir o sobrevivir. En estos casos modela el logaritmo de una proporción (p), que bien podría ser la incidencia o la prevalencia de una enfermedad.⁽³⁾
- Regresión de Poisson: empleada cuando la variable dependiente se ajusta bien a una distribución de Poisson (distribución que representa la probabilidad de que un determinado número de eventos ocurra durante un periodo de tiempo en un espacio o población especificada), para cualquier combinación de valores de la covariable. Una variable Poisson

se presentan como la tasa de un evento de interés, por ejemplo, el número de fallecidos por determinada causa en una población específica durante el periodo de observación. Esta técnica modeliza bien situaciones de conteo en una unidad de tiempo.^(9,10)

- Regresión discontinua: apropiada en diseños cuasi-experimentales pretest-postest en los que se investiga el efecto de las intervenciones asignadas por encima o por debajo de un valor de corte o umbral. Mediante la comparación de las observaciones en un entorno del valor umbral, es posible estimar el valor del efecto de tratamiento promedio en los casos que la aleatorización es imposible.^(11,12)

Para hacer un uso correcto de los resultados de una regresión se debe conocer el papel que cumplen las covariables, su adecuada codificación e inclusión en el modelo seleccionado, los supuestos básicos de la regresión, así como los métodos para evaluar el ajuste del modelo y la interpretación de los resultados.

En general, en el caso de desenlaces continuos, se pueden generar modelos de regresión lineal múltiple o modelos más flexibles como los aditivos generalizados que transforman automáticamente las covariables que no tienen una relación lineal con el desenlace. En el caso de desenlaces dicotómicos, se generan modelos de regresión logística o extensiones como la regresión logística bayesiana naive, los modelos aditivos generalizados, las particiones recursivas de un árbol de regresión o de clasificación u otros modelos de inteligencia artificial como las redes neuronales y las máquinas de soporte vectorial. Para desenlaces en función del tiempo hasta el evento se usa el método de riesgos proporcionales de Cox o métodos paramétricos que modelan en relación con las distribuciones exponencial, Weibul o de Poisson, entre otras⁽⁸⁾

Para validar estos modelos se puede dividir la muestra en varias partes en cada una de las cuales el modelo se vuelve a estimar y se comparan los resultados para lo cual se pueden emplear las técnicas de remuestreo, donde se determina que el modelo es válido si la predicción del desenlace es igual en los datos en que se desarrolló y en los datos nuevos.

Las técnicas de remuestreo como el *jackknife* o el *bootstrap* consisten en seleccionar n muestras (submuestras) a partir de una muestra inicial. Esa selección de muestras se hace con reemplazo, con el empleo de la simulación computacional. El método de remuestreo es una buena opción para realizar estimaciones cuando los estadísticos son difíciles de manipular algebraicamente. El método *Jackknife* es una aproximación lineal del *bootstrap*, útil para la varianza y el sesgo.⁽¹³⁾

Los problemas en la no validez de estos modelos pueden deberse a deficiencias en el diseño, sobreajuste a los datos de la cohorte inicial, deficiencias o diferencias en el sistema de salud, diferencias en los métodos de medición y en las características de los pacientes.⁽⁸⁾

Ecuaciones estructurales

Los modelos de ecuaciones estructurales surgen para flexibilizar los modelos de regresión, al ser menos restrictivos al incluir errores de medida en las variables dependientes e independientes. Matemáticamente, son más complejos de estimar y por ello su uso no se extendió hasta 1973, cuando apareció el programa LISREL;⁽¹⁴⁾ ya en 2007 se facilitó la estimación en entorno gráfico, a través del programa AMOS.⁽¹⁵⁾

Estos modelos comprenden una serie de procedimientos que permiten probar la relación entre variables observadas (medidas directamente) y latentes (constructos medidos a través de otras variables observadas).^(16,17)

Cuando el modelo se compone solo de variables observadas, la técnica se conoce como análisis de ruta, de camino o de sendero (*path analysis*) y es similar al análisis de regresión lineal clásico con la ventaja de que es posible estimar el efecto indirecto y total que tiene una variable sobre otra.⁽¹⁸⁾

Cuando el modelo mezcla ambos tipos de variables se convierte en análisis de estructura de covarianza. Este análisis puede ser recursivo si admite una dependencia unidireccional o no recursivo si se da una relación recíproca entre dos variables.⁽¹⁾ Además puede ser un análisis factorial confirmatorio para estimar la correlación entre las variables latentes. En cualquier caso, se trata de verificar cuanto ajustan los datos a partir de un modelo teórico de supuestas relaciones causales propuesto por el investigador.

Si el modelo obtenido indica que los datos no respaldan la teoría propuesta por el investigador, habrá que relaborar el modelo teórico para una nueva comprobación. También es necesario verificar el cumplimiento de supuestos como linealidad, normalidad, entre otros y que existan entre 15 y 20 casos por cada variable medida.⁽¹⁸⁾ Es posible verificar con estas ecuaciones un modelo de regresión logística obtenido a partir de un estudio de casos y controles para la identificación de factores de riesgo o pronóstico como el estudio de Alves de Oliveira.⁽¹⁹⁾

Ruiz⁽²⁰⁾ plantea que son muchos los tipos de modelos que se pueden definir con esta metodología, entre ellos menciona la regresión múltiple con multicolinealidad, el análisis factorial confirmatorio, el análisis factorial de segundo orden, el análisis de sendero, el modelo causal completo con variables latentes, el modelo de curva latente, el análisis de mediación y los modelos multinivel, multigrupo y los basados en las medias como ANOVA, ANCOVA, MANOVA y MANCOVA.

Los puntos fuertes de estos modelos están en haber desarrollado unas convenciones que permiten su representación gráfica, la posibilidad de hipotetizar efectos causales entre las variables, la concatenación de efectos y las relaciones recíprocas entre variables.⁽¹⁹⁾

Análisis estadístico implicativo

El análisis estadístico implicativo es otra técnica estadística multivariadas de la minería de datos basada en la teoría de la cuasi-implicación, la inteligencia artificial y el álgebra booleana, para modelar la cuasi-implicación entre los sucesos y variables de un conjunto de datos.

Esta técnica fue creada por el francés Régis Gras ^(21,22,23) para solucionar problemas de la Didáctica de las Matemáticas. La misma contempla la estructuración de datos, interrelaciona sujetos y variables, la extracción de reglas inductivas entre las variables y, a partir de la contingencia de estas reglas, la explicación y en consecuencia una determinada previsión en distintos campos del saber.

Con esta técnica se estiman las frecuencias absolutas de ocurrencia de cada variable, sus medias y desviaciones estándares, las frecuencias de ocurrencias de cada par de variables que se pueden formar, así como sus coeficientes de correlación, los índices de similaridad, de cohesión implicativa y de implicación inclusión, las tipicalidades y contribuciones de los individuos. Los resultados se presentan en tres gráficos que facilitan la interpretación de los mismos: los árboles de similaridad y de cohesión y el grafo implicativo que puede ser general y también en modo cono, según decida el investigador. Si se fuera a ubicar esta técnica dentro de la clasificación de las técnicas empleadas con regularidad, (Figura 1) se diría que es una técnica mixta.

Varias investigaciones recientes han identificado las ventajas de esta técnica y han validado su efectividad en la identificación de factores pronósticos y han verificado una utilidad mayor al emplearse previo a la regresión logística ya que aporta otro criterio para decidir cuales covariables incluir en el modelo de regresión.^(24,25,26,27,28,29,30,31,32)

Análisis multivariado espacial

La dimensión espacial también es relevante, en los estudios de la causalidad en Ciencias Médicas, para resumir estructuras de covarianza e identificar patrones espaciales, lo cual constituye un desafío desde el punto de vista estadístico. Unas variedades de métodos han sido desarrolladas para incorporar la información espacial como el basado en índices de autocorrelación espacial, el semivariograma a partir de una variable sintética resumen, los mapas sintéticos 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, entre otros.⁽³³⁾

Los métodos gráficos multivariados fueron desarrollados por Sagaró y Zamora,⁽³⁴⁾ de igual forma los métodos avanzados que se emplean de conjunto con las técnicas multivariadas para garantizar la validez de los estudios observacionales de causalidad en salud fueron abordados por estas autoras en otro artículo, así como otros detalles en relación al surgimiento de estas técnicas.^(35,36)

CONCLUSIONES

Aunque las técnicas estadísticas multivariadas más empleadas en el estudio de la causalidad son las explicativas, como la regresión y los modelos estructurales, a veces es necesario un primer análisis exploratorio donde se hace necesario el empleo de las técnicas multivariadas descriptivas, por lo que es importante conocer todas las técnicas y en qué momento emplearlas. Se propone el análisis estadístico implicativo como técnica multivariada de uso habitual en las investigaciones médicas para la determinación de factores de riesgo y pronósticos por sus bondades descriptivas, explicativas y predictivas, además de sus prestaciones gráficas, que permiten visualizar las relaciones entre un conjunto de variables.

Conflicto de intereses

Los autores declaran que no existe conflicto de intereses.

Contribución de los autores

Todos los autores contribuyeron en igual medida a la realización de la investigación.

Financiación

Los autores declaran que no recibieron ningún tipo de financiación para el desarrollo de la presente investigación.

REFERENCIAS BIBLIOGRÁFICAS

1. Hair JF, Anderson RE, Tatham RL, Black WC. Análisis multivariante 5ª. Ed. [Internet] Prentice Hall International. Inc; 1999. [citado 10/01/2018]. Disponible en: <https://www.scribd.com/document/367039653/Analisis-Multivariante-5ta-Edicion-Joseph-F-Hair-LIBROSVIRTUAL-com>

2. Hernández Avila M, Garrido Latorre F. Diseño de estudios epidemiológicos. Salud Pública de México [Internet] 2000 marzo-abril [citado 10/01/2018]; 42(2): [aprox. 10p.]. Disponible en: <https://www.scielosp.org/pdf/spm/2000.v42n2/144-154>
3. Diaz-Quijano FA. Regresiones aplicadas al estudio de eventos discretos en epidemiología. Rev Univ Ind Santander Salud [Internet] 2016 [citado 18/03/2019]; 48(1): [aprox. 6p.]. Disponible en: http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S0121-08072016000100003
4. Lozares Colina C, López-Roldán P. El análisis multivariado: definición, criterios y clasificación. Revista de Sociología [Internet] 1991 [citado 18/03/2019]; (37): [aprox. 21p.]. Disponible en: <https://papers.uab.cat/article/view/v37-lozares-lopez>
5. Closas AH, Arriola EA, Kuc CI, Amarilla MR, Jovanovich EC. Análisis multivariante, conceptos y aplicaciones en Psicología Educativa y Psicometría. Enfoques [Internet] 2013 [citado 18/03/2019]; 25(1): [aprox. 20p.]. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=5229555>
6. Olive DJ. Robust Multivariate Analysis. [Internet] Switzerland: Springer International Publishing AG; 2017. [citado 18/03/2019] Disponible en: <https://link.springer.com/content/pdf/10.1007/978-3-319-68253-2.pdf>
7. Ranganathan P, Pramesh CS, Aggarwai R. Common pitfalls in statistical analysis: Logistic regression. Perspect Clin Res [Internet] 2017 [citado 15/01/2019]; 8(3): [aprox. 4p.]. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5543767/>
8. Isaza-Jaramillo S, Jaimes-Barragán F. Ronda clínica y epidemiológica: aproximación a los modelos de predicción clínica. Iatreia [Internet]. 2017 [citado 15/01/2019]; 30(1): [Aprox. 7p.]. Disponible en: http://www.scielo.org.co/scielo.php?pid=S0121-07932017000100092&script=sci_abstract&tlng=es
9. Gentzkow M, Shapiro JM, Taddy M. Measuring Polarization in High-Dimensional Data: Method and Application to Congressional Speech. SIEPR Discussion Paper No. 16-028 Stanford Institute for Economic Policy Research; 2016. [citado 15/01/2019]. Disponible en: <https://siepr.stanford.edu/sites/default/files/publications/16-028.pdf>
10. Bind MA. Causal Modeling in Environmental Health. Annu. Rev. Public Health [Internet] 2019 [citado 15/01/2019]; 40: [aprox. 9p.]. Disponible en: <https://www.annualreviews.org/doi/abs/10.1146/annurev-publhealth-040218-044048>
11. Lee L. Regression Discontinuity Designs in Economics. Journal of Economic Literature [Internet] 2010 [citado 15/01/2019]; 48(2): [aprox. 20p.]. Disponible en: <http://dx.doi.org/10.1257%2Fjel.48.2.281>
12. Ponce C, Escobal J. Metodología cuantitativa de la evaluación del impacto. En: Escobal J, Ponce C. (Eds.) Combinando protección social con generación de oportunidades económicas: una evaluación de los avances del programa Haku Wiñay [Internet]. Lima: GRADE Group for the Analysis of Development; 2016. [citado 20/01/2019]. Disponible en: <https://www.grade.org.pe/publicaciones/metodologia-cuantitativa-de-la-evaluacion-del-impacto/>

13. Monsalve Torra AE. Sistema de ayuda a la decisión clínica en enfermedades de diagnóstico complejo. [Tesis Doctoral]. Universidad de Alicante; 2017 [citado 21/02/2019]. Disponible en: https://rua.ua.es/dspace/bitstream/10045/65334/1/tesis_monsalve_torra.pdf
14. Jöreskog KG, Olsson UH, Wallentin FY. Multivariate Analysis with LISREL. [Internet] Switzerland: Springer International Publishing; 2016. [citado 21/02/2019]. Disponible en: <https://link.springer.com/content/pdf/10.1007/978-3-319-33153-9.pdf>
15. Roth Unzueta E. Análisis multivariado en la investigación psicológica: Modelado predictivo y causal en SPSS y AMOS. [Internet] Bolivia: SOIPA Ltda; 2012. [citado 21/02/2019]. Disponible en: <http://repositorio.ucb.edu.bo/xmlui/bitstream/handle/UCB/134/LibroAnlisisMultivariadopescaneadonuevo-comprimido.pdf?sequence=1&isAllowed=y>
16. Manzano Patiño AP. Introducción a los modelos de ecuaciones estructurales. Inv Ed Med [Internet]. 2017 [citado 18/01/2019]; 7(25): [aprox. 5p.]. Disponible en: <http://www.scielo.org.mx/pdf/iem/v7n25/2007-5057-iem-7-25-67.pdf>
17. Escobedo Portillo MT, Hernández Gómez JA, Estebané Ortega V, Martínez Moreno G. Modelos de Ecuaciones Estructurales: Características, Fases, Construcción, Aplicación y Resultados. Cienc Trab. [Internet] 2016 [citado 18/01/2019]; 18(55): [aprox. 6p.]. Disponible en: https://scielo.conicyt.cl/scielo.php?script=sci_abstract&pid=S0718-24492016000100004&lng=es&nrm=iso
18. Keith TZ. Multiple regression and beyond: An introduction to multiple regression and structural equation modeling. 2nd ed. [Internet]. New York: Taylor & Francis; 2015. [citado 18/01/2019]. Disponible en: <http://admin.ifj.org/multiple-regression-and-beyond-an-introduction-to-multiple-regression-and-structural-equation-modeling.pdf>
19. Alves de Oliveira A, Furquim de Almeida M, Pereira da Silva Z, Lisiane de Assunção P, Rigo Silva AM, Geremias dos Santos H, et al. Factors associated with preterm birth: from logistic regression to structural equation modeling. Cad. Saúde Pública [Internet]. 2019 [citado 18/01/2019]; 35(1): [aprox. 2p.]. Disponible en: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2019000104001&lng=en&nrm=iso&tlng=en
20. Ruiz M, Pardo A, San Martín R. Modelos de ecuaciones estructurales. Papeles del Psicólogo [Internet] 2010 [citado 20/01/2019]; 31(1): [aprox. 11p.]. Disponible en: <http://www.redalyc.org/pdf/778/77812441004.pdf>
21. Gras R, Régnier JC, Lahanier-Reuter D, Marinica C, Guillet F. L'Analyse Statistique Implicative. Des Sciences dures aux Sciences Humaines et Sociales 3ème édition revue et augmentée. Cépaduès Ed; 2017.
22. Gras R, Bodin A. L'A.S.I., analyseur et revelateur: de la complexité cognitive taxonomique. EMP [Internet]. 2017 [citado 22/01/2019]; 19(1): [aprox. 18p.]. Disponible en: <http://dx.doi.org/10.23925/1983-3156.2017v19i1p497-515>
23. Bodin A, Giovannini ML, Silva L. L'utilisation de l'analyse implicative et cohésitive pour exploiter un test standardisé de mathématique. Italian Journal of Educational Research [Internet]. 2017 [citado 20/01/2019]; 19. Disponible en: <https://ojs.pensamultimedia.it/index.php/sird/article/view/2547>

24. Sagaró Del Campo NM, Zamora Matamoros L. ¿Por qué emplear el análisis estadístico implicativo en los estudios de causalidad en salud? RCIM [Internet]. 2019 [citado 20/06/2019]; 11(1): [aprox. 15p.]. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1684-18592019000100088
25. Sagaró Del Campo N, Zamora Matamoros L. Análisis estadístico implicativo versus Regresión logística binaria para el estudio de la causalidad en salud. MULTIMED [Internet] 2019 [citado 19/12/2019]; 23(6): [aprox. 24p.]. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1028-48182019000601416&lng=es
26. 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 [Internet]. 2015 [citado 20/01/2018]; 19(8). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192015000800003
27. 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 [Internet]. 2016 [citado 20/01/2018]; 20(3). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192016000300010
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 mama. MEDISAN [Internet]. 2017 [citado 20/01/2018]; 21(4): [aprox. 1p.]. Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192017000400003
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 próstata. MEDISAN [Internet]. 2018 [citado 20/01/2018]; 22(1). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192018000100007
30. 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 [Internet]. 2018 [citado 20/01/2019]; 22(6). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192018000600007
31. Galano Vázquez K, Sagaró del Campo NM, Zamora Matamoros L, Lambert Matos Y, Mingui Carbonell E. Análisis estadístico implicativo en la identificación de factores pronósticos de mortalidad del cáncer renal. Rev. inf. Cient [Internet]. 2018 [citado 14/05/2019]; 98(2). Disponible en: <http://www.revinfcientifica.sld.cu/index.php/ric/article/view/2268>
32. Pardo-Santana S, Sagaró-del-Campo NM, Zamora-Matamoros L, Viltre-Castellanos DM. Utilidad del análisis estadístico implicativo para identificar factores pronósticos en pacientes con cáncer de mama. Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta [Internet]. 2019 [citado 25/05/2019]; 44(4). Disponible en: <http://revzoilomarinellosld.cu/index.php/zmv/article/view/1869>
33. Balzarini M, Bruno C, Córdoba M, Teich I. Herramientas en el Análisis Estadístico Multivariado. Escuela Virtual Internacional CAVILA Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba, Argentina; 2015.

-
34. Sagaró Del Campo NM, Zamora Matamoros L. Métodos gráficos en la investigación biomédica de causalidad. Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta [Internet]. 2019 [citado 12/06/2019]; 44(4). Disponible en: <http://revzoilomarinello.sld.cu/index.php/zmv/article/view/1846>
35. Sagaró Del Campo NM, Zamora Matamoros L. Métodos actuales para asegurar la validez de los estudios de causalidad en medicina. Gac. Méd. Espirit [Internet]. 2019 [citado 20/10/2019]; 21(2): [aprox. 14p.]. Disponible en: <http://scielo.sld.cu/pdf/gme/v21n2/1608-8921-gme-21-02-146.pdf>
36. Sagaró Del Campo NM, Zamora Matamoros L. Evolución histórica de las técnicas estadísticas y las metodologías para el estudio de la causalidad en ciencias médicas. MEDISAN [Internet]. 2019 [citado 11/06/2019]; 23(3). Disponible en: http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1029-30192019000300534

ANEXO. Terminología de uso frecuente

- **Modelo:** Representación abstracta y simplificada de un proceso que afecta a una población. Generalmente esa representación es de tipo matemático y su objetivo consiste en analizar y predecir los procesos de forma cuantitativa. Puede construirse bajo condiciones experimentales o naturales
- **Modelo de regresión:** Representación matemática de la relación entre una variable dependiente o de respuesta
- y una combinación de variables independientes (llamadas también variables explicativas, predictoras o covariables).
- **Individuos:** Elementos sobre los cuales se observa una o más características de interés.
- **Variable:** representación numérica de una característica de los individuos que componen la muestra que toma un valor diferente para cada uno de ellos. Ejemplo: Edad, sexo, escolaridad, peso, talla
- **Variable respuesta o dependiente:** Variable objeto del estudio cuyos cambios se pretenden explicar por medio de las variables explicativas o independientes.
- **Variables independientes o explicativas:** Variables que permiten construir un modelo que explica el comportamiento de una o más variables respuesta.
- **Variable de confusión o confusora:** Variable que está simultánea e independientemente asociada al factor de riesgo y a la enfermedad estudiada y puede modificar el resultado del estudio. Cuando se debe a un sesgo en el proceso de asignación, el error resultante se denomina sesgo de selección
- **Validez:** Solidez o rigor de un estudio en relación con el grado de aproximación a la "verdad" de sus resultados. Un estudio es válido si el modo en que ha sido diseñado y realizado hace que los resultados no estén sesgados.
- **Validez interna:** Grado en el que los efectos observados son verdaderos para las personas del estudio.
- **Validez externa:** Grado en que los resultados de un estudio son generalizables a otros individuos.
- **Estimación:** Obtención de conocimiento sobre cierto aspecto de la población desconocido a partir de lo encontrado en una muestra.
- **Ajuste:** Conjunto de técnicas que se emplean después de la recogida de datos para controlar o tener en cuenta el efecto de la variable de confusión, sean conocidas o potenciales
- **Robustez:** Propiedad de una prueba estadística que permite violar sus supuestos sin que ello repercuta sustancialmente en las conclusiones. En el caso de los modelos robustos no se afectan significativamente cuando se introducen cambios en él, lo cual se corrobora con un análisis de sensibilidad.
- **Homocedasticidad:** Igualdad de las varianzas poblacionales en los distintos grupos de estudio.
- **Valor esperado:** valor central o medio que se corresponde con la mayor cantidad de valores de la variable. Ejemplo: el promedio aritmético