

USOS Y LIMITACIONES DE LOS MÉTODOS DE ANÁLISIS MULTIVARIADOS EN LA INVESTIGACIÓN EPIDEMIOLÓGICA¹

Hugo Grisales Romero*

Fecha de recibo: Marzo/06

Fecha aprobación: Julio/06

El análisis multivariado comprende un grupo de técnicas estadísticas, en las cuales se analiza la aleatoriedad de múltiples variables al mismo tiempo. Un concepto unificador que subyace en las diferentes técnicas y facilita el análisis multivariado, es el uso de modelos matemáticos que responden de la interrelación entre las variables del análisis¹. La planificación de una gran parte de la investigación en salud pública en general y en epidemiología en particular, requiere cada vez más la aplicación de las técnicas que provee el análisis multivariado, las cuales contribuyen a la comprensión del(los) evento(s) de interés desde una perspectiva multifactorial².

Los fundamentos teóricos del análisis multivariado se desarrollaron durante los tres primeros decenios del siglo XX por Fisher y Pearson y su primera disciplina de aplicabilidad fue en la agricultura, siendo el análisis discriminante la técnica más utilizada, la cual, a través de los años ha dado paso a otras técnicas menos exigentes en supuestos. Sólo a partir

de la segunda guerra mundial se amplió el horizonte aplicativo a diversas disciplinas, entre ellas la epidemiología donde Cornfield, Truett y Kleinbaum contribuyeron en la aplicación de las técnicas multivariadas en sus estudios sobre factores de riesgo para enfermedad coronaria³.

Usos

Los métodos estadísticos multivariados pueden agruparse en dos conjuntos: los que permiten extraer la información acerca de la interdependencia entre las variables que caracterizan cada uno de los individuos, y los que permiten extraer información acerca de la dependencia entre una (o varias) variable(s) con otra (u otras)⁴. Del primer grupo referido hacen parte el análisis factorial, de clusters, de correlación canónica, de ordenamiento multidimensional y de correspondencias; en el segundo grupo se encuentran las técnicas de regresión, los análisis de contingencia múltiple, el análisis de la covarianza y el análisis discriminante.

* Profesor Facultad Nacional de Salud Pública. Universidad de Antioquia

¹ Algunos autores como Kleinbaum no hablan de métodos multivariados si no de métodos multivariables, donde enfatiza es en la evaluación simultánea del efecto de varias variables independientes sobre una variable respuesta (análisis de dependencia).

En los últimos tiempos, en la investigación epidemiológica de tipo analítico, el análisis factorial, como técnica de interdependencia y los modelos de regresión logística, de Cox y de Poisson como herramientas de dependencia, han tenido un papel protagónico como métodos de análisis, lo cual podría explicarse por la tecnología informática que ha proporcionado el software para facilitar las elaboraciones. Lo claro es que las técnicas antes indicadas han robustecido el arsenal de herramientas que tiene el epidemiólogo para abordar los objetivos de sus investigaciones.

El análisis factorial es un método de interdependencia, que sirve para determinar el número y la naturaleza de un grupo de constructos-atributos para explicar un fenómeno que está subyacente en un conjunto de variables. Por ejemplo, un epidemiólogo podría estar interesado en establecer los factores que inciden en la satisfacción laboral en los trabajadores sanitarios de un municipio o también de establecer los dominios que permitan determinar las condiciones de vida de los adolescentes de una región. En estos ejemplos, el análisis factorial desempeñaría un papel fundamental para resolver los interrogantes en cada caso.

La regresión de riesgos proporcionales de Cox es un método de dependencia, utilizado en estudios longitudinales sobre datos de supervivencia y donde la función de riesgo da la probabilidad que un paciente muera durante un intervalo de tiempo específico, dado que el paciente vivió hasta el principio del mismo; en este modelo, las covariables tienen un efecto multiplicativo o proporcional en la probabilidad de morir; de aquí el nombre de riesgos proporcionales.

Por ejemplo, mediante la construcción de un modelo de Cox se puede contrastar la hipótesis sobre los efectos del consumo de alcohol y el sexo hasta el momento del comienzo de la cirrosis hepática. En esta línea, se podría responder: ¿tienen los hombres y las mujeres diferentes riesgos de desarrollar cirrosis hepática a causa del consumo de alcohol?⁵

La regresión logística es quizás la técnica de dependencia más utilizada en la investigación epidemiológica de tipo analítico, para predecir o ajustar el comportamiento de una variable dependiente dicotómica, por lo general la enfermedad o algún otro evento recurrente en función de otras variables llamadas covariables. Por ejemplo, mediante esta técnica es factible predecir las variables que explican mejor el bajo peso al nacer en un estudio analítico.

La regresión de Poisson es una técnica de dependencia donde la variable dependiente hace referencia a años-persona-riesgo-exposición como es el caso de las tasas de incidencia, y que corresponden a un evento que se considera de baja frecuencia o 'raro'. Por ejemplo, un epidemiólogo podría estar interesado en comparar la incidencia de cáncer de piel en dos ciudades, considerando que las diferencias en las incidencias puedan ser debidas a que ambas ciudades tengan una distinta pirámide de población o quizás, y sería una hipótesis más interesante a investigar, la presencia de algún otro factor.

Finalmente, si se conoce la distribución de las poblaciones para los distintos grupos de edad, así como el grupo al que pertenece cada enfermo, se puede plantear un modelo de Poisson.

En el último decenio en investigación epidemiológica, se ha pretendido incorporar el contexto en el cual suceden los eventos y las exposiciones y se conoce como eco epidemiología. Quizás la herramienta de dependencia más potente es el análisis multinivel, que como técnica permite diferenciar entre efectos individuales y los efectos de contexto. Por ejemplo un epidemiólogo podría estar interesado en estudiar a nivel regional las condiciones de vida de los adolescentes jóvenes y plantear en alguno de sus objetivos una comparación de dichas condiciones, según características individuales y de contexto. En este caso se puede establecer una estructura jerárquica de 3 niveles, donde los adolescentes jóvenes se sitúan en el nivel 1, la familia en el nivel 2 y la comuna en el nivel 3. Otras aplicaciones de esta técnica en epidemiología se centran en la evaluación de indicadores de salud que se distribuyen por áreas; en la satisfacción de los usuarios con relación a especialistas que los asisten y el centro de salud donde se atienden; en pruebas clínicas donde se aplican diferentes tratamientos, y en la evolución de la salud de un paciente en el tiempo, entre otros.

La necesidad de aplicar metodologías más allá de las puramente individualistas ha sido reconocida por diferentes autores, y se refleja en las recientes publicaciones de textos sobre análisis multinivel en salud⁶.

Limitaciones

La eficiencia estadística de los modelos multivariados tiene un precio, que es la suposición que una forma matemática dada describa perfectamente la relación de las variables del estudio. Esta suposición es peligrosa ya que en el paradigma de la multicausalidad, es una utopía

pensar que se han controlado todas las variables que conforman el diagrama causal.

La confusión residual o ajuste incompleto se presenta cuando éste no elimina completamente el efecto de confusión debido a una variable dada o a un conjunto de variables. Ocurre cuando algunas categorías del confusor controlado son tan amplias que se obtiene un ajuste imperfecto o cuando algunas variables de confusión no se controlan. Entre las principales causas de esta limitante en las técnicas de dependencia se destacan: la definición inadecuada de las categorías de la(s) variable(s) de confusión, la variable empleada para el ajuste como un sucedáneo imperfecto de la condición o de la característica por la cual se desea ajustar, otros importantes confusores no fueron incluidos en el modelo o errores de clasificación en las variable de confusión⁷.

El sobreajuste, el cual ocurre cuando se ajusta erróneamente para una variable que o está en el camino causal que vincula la exposición con el desenlace-causa intermedia, o tiene una relación tan fuerte con la exposición o con el desenlace, que la verdadera relación se distorsiona, es otra limitante del modelo de dependencia, el cual exige la necesidad de considerar los determinantes biológicos de la relación que se valora, evaluar las relaciones entre las variables de confusión postuladas por una parte y la exposición y las variables de desenlace por otra.

El problema de qué variables independientes confusoras han de ser incluidas en el modelo, está en el centro de discusión sobre el tópico de confusión y de confusión residual⁸

Se obvia, cuando se construyen modelos de dependencia en la realización del análisis de sensibilidad del modelo, esto es examinar si se obtienen resultados similares cuando se han empleado diferentes modelos o se han hecho diferentes supuestos para el análisis.

Es aconsejable no utilizar mecánicamente los procedimientos de selección paso a paso, cuando se construyen modelos de regresión. Éstos son automáticos, tienen una alta probabilidad de generar hallazgos espurios y en muchas ocasiones las variables que son retenidas en el modelo, clínicamente no son importantes, en contraposición con aquellas que se excluyeron.

Una advertencia importante es no ser informal en la aplicación de las técnicas, sin revisión del cumplimiento de los supuestos que las sustentan. De hecho, técnicas como la regresión múltiple exige el cumplimiento de algunos supuestos que en la mayoría de las situaciones se omiten, lo cual induce a decisiones derivadas de modelos mal ajustados.

No se deben utilizar las técnicas multivariadas por moda sino que su uso debe estar en consonancia directa con los objetivos trazados; por lo tanto, si los objetivos no exigen el uso de la técnica, es un exabrupto utilizarla por el mero hecho de darle formalidad teórica y sofisticación al estudio.

¿Cuál de los modelos a elegir es el más apropiado? Al respecto, tal como lo sugiere Kleinbaum⁹, hay varios aspectos que el investigador debe atender previamente antes de elegir el 'mejor' modelo: ¿Cómo identificar los roles que juegan los diferentes predictores? ¿Qué estrategias de modelación se seguirán? ¿Cómo se considerarán varias exposiciones y varias variables de control? ¿Cómo estimar la confusión si la interacción está presente? ¿Cómo estimar colinealidad y mejor aún, cómo controlarla? Definitivamente es un contrasentido, aunque común, la construcción automática de modelos multivariados en los cuales no se tiene claridad del propósito que se persigue con el análisis o se desconocen aspectos importantes del método. Este es el abrebocas para cometer serias equivocaciones.

REFERENCIAS BIBLIOGRÁFICAS

- 1 Rothman KJ. *Epidemiología Moderna*. Ediciones Díaz de Santos. 1986. p. 318.
- 2 Johnson RA, Wichern DW. *Applied multivariate statistical analysis*. Prentice Hall, 1982; 461-517.
- 3 Londoño, J.L. *Metodología de la investigación epidemiológica*. 3ª edición. Manual Moderno. 2004. p. 178
- 4 Pla, Laura E. *Análisis Multivariado: Método de componentes principales*. Monografía preparada para su publicación en el Departamento de Asuntos Científicos y Tecnológicos de la Secretaría General de la OEA. Editora: Eva V. Chesneau. Washington, 1986. p. 5.
- 5 Beth Dawson-Saunders, Robert G. Trapp. *Bioestadística Médica*. Editorial el manual Moderno, S.A. de C.V. México, D.F. p. 240.
- 6 Catalán-Reyes MJ, et al. *Utilización de los modelos multinivel en investigación sanitaria*. Gaceta Sanitaria 2003; 17(Supl 3):35-52.
- 7 Szklo M., Nieto J. *Epidemiología Intermedia*. Conceptos y Aplicaciones. Ediciones Díaz de Santos. 2003. p. 285 a 293
- 8 Greenland S. *Modeling and variable selection in epidemiologic analysis*. Am J Public Health 1989; 79: 340-349
- 9 Kleinbaum DG. *Epidemiologic methods: the «art» in the state of the art*. Journal of Clinical Epidemiology. 55 (2002) 1196-1200. p. 1198.