

El uso de las técnicas de análisis multivariado en la investigación médica mexicana

Efrén Raúl Ponce Rosas, Ma. Eloísa Dickinson Bannack, Francisco J.F. Gómez Clavelina
Departamento de Medicina Familiar, Facultad de Medicina, UNAM

(Recibido, febrero 7, 1995; aceptado, julio 3, 1995)

La estadística moderna es reconocida como una herramienta útil para la investigación médica en general; en la actualidad, los comités editoriales de las revistas de mayor prestigio informan con más frecuencia a los autores de artículos médicos sobre el tipo de información que debe proporcionarse de los métodos estadísticos utilizados en sus trabajos.

En 1988, el Comité Internacional de Editores de Revistas Médicas (CIERM)¹ publica quince lineamientos básicos para presentar por escrito los aspectos estadísticos de la investigación realizada, de tal forma que responda a las solicitudes de los editores y revisores y que resulte útil a los lectores; los quince lineamientos mencionados se adicionaron a los del artículo original² sobre los requisitos uniformes para preparar manuscritos que se proponen para publicación de revistas médicas que el CIERM representa. Aunque el grupo del CIERM publicó sus lineamientos por primera vez en 1979, han existido diversas enmiendas^{3,4} e incluso existen versiones en castellano, una se publicó en la serie Comunicación Biomédica del boletín de la Oficina Sanitaria Panamericana en 1989⁵ (en la cual se afiliaron más de 300 revistas médicas), y otra en nuestro país editada en la Revista de Salud Pública de México en 1992⁶.

Los quince lineamientos propuestos⁷ se refieren a:

- La descripción detallada y el porqué de los métodos estadísticos elegidos que conduzcan a valorar la confiabilidad de los datos y la validez de los resultados.
- La cuantificación de resultados medidos con otra a través de intervalos de confianza.
- La no dependencia exclusiva de los valores de probabilidad (p).

- Análisis de la elegibilidad de los sujetos estudiados.
- Descripción detallada del proceso de aleatorización.
- El método de enmascaramiento de las observaciones (método ciego).
- Información de las complicaciones del tratamiento.
- Especificación del número de observaciones (tamaño de la muestra).
- Causas y motivos de las pérdidas de sujetos observados.
- Referencias actualizadas en relación al diseño del estudio y métodos estadísticos de trabajo.
- Especificación de programas de computación utilizados.
- Información del método estadístico empleado en el análisis y su ubicación en el texto propuesto a publicación.
- Uso al mínimo de cuadros y figuras sin repetición de datos y resultados.
- Evitar el uso no técnico de términos estadísticos.
- Definición de términos, abreviaturas y símbolos.

Además de los lineamientos señalados, es importante considerar otros como:

- El diseño estadístico de investigación (comparaciones independientes o pareadas, número de grupos analizados y número de mediciones por unidad de muestreo).

- Las variables, su naturaleza y escala de medición, así como la distribución de sus mediciones (normal, binomial, hipergeométrica, etcétera.) y su posible transformación para obtener "normalidad".
- Elección adecuada de la prueba de hipótesis seleccionada (paramétrica o no paramétrica), acorde a la distribución de los datos y el diseño estadístico empleado.

Todos estos términos, en la medida en que son correctamente concebidos, utilizados, interpretados y expuestos con claridad en las investigaciones generadas en el campo de la medicina, enriquecen y fortalecen la calidad de las mismas en función de que describen la información pertinente y necesaria de un trabajo de investigación para poder ser analizado y evaluado de una manera crítica.

En los últimos años, el uso de las técnicas estadísticas de análisis multivariado (TEAM) se ha adicionado a la investigación y debe ser considerado y conceptualizado en función de la importancia que tiene en el diseño, conducción, análisis y evaluación de los trabajos de investigación.

Todos los investigadores describen e interpretan los resultados de una sola variable: para la gran mayoría, analizar dos variables simultáneamente les resulta cotidiano, solamente se necesita saber si se desea comparar (promedios o proporciones) o asociar (correlación, regresión o análisis de asociación mediante tablas de contingencia) entre dos variables.

¿Qué sucede si se desea comparar o asociar entre tres o más variables simultáneamente?; con tres variables resulta relativamente sencillo representarlas en un plano cartesiano tridimensional (ejes X, Y, Z), sin embargo, representar gráficamente (con todas las ventajas que un gráfico posee), un plano multidimensional de cuatro, cinco, diez o más ejes, solamente es concebido en la subrealidad; éste es el campo de las TEAM, en el cual la subrealidad mencionada produce serios problemas de diseño, análisis, interpretación y evaluación integral de trabajos de investigación.

Para resolver este problema, las TEAM disminuyen la multidimensionalidad mediante una representación numérica y gráfica simple y de más fácil conceptualización y que permita al lector de revistas médicas comprender mejor los resultados sobre una base más objetiva y racional.

Las TEAM utilizan para representar gráficamente los resultados de su aplicación, la estructura conocida como la más sencilla y fácilmente interpretable: el plano cartesiano bidimensional. Es importante mencionar que la multidimensionalidad se reduce a dos dimensiones mediante técnicas de análisis de las propias TEAM. Ejemplo de ello en nuestro medio, son algunas publicaciones recientes que han informado resultados de investigaciones en las cuales se han usado dichas técnicas.

Diversos autores⁸⁻¹³ han utilizado diferentes técnicas de análisis multivariado en investigaciones realizadas en el área de la salud pública, encontrándose al análisis factorial como la técnica más frecuente: dicha técnica también fue utilizada en otro trabajo realizado por los autores del presente¹.

En el análisis detallado de las referencias mencionadas se encuentra que algunos autores no proporcionan información suficiente sobre las técnicas utilizadas, otros dan información que resulta difícil comprender si no se conoce sobre la teoría de las TEAM, y son escasos los que dan información clara y precisa.

Aunque existen diversas TEAM (análisis factorial, componentes principales, análisis de conglomerados, análisis discriminante, regresión lineal multivariada, regresión logística multivariada, correlación múltiple, correlación canónica y análisis de varianza multivariada entre otros), en los trabajos revisados el más frecuente fue el análisis factorial (análisis de factores, "factor analysis"). Dicha técnica tiene como objetivo explicar las relaciones entre tres o más variables², y se calcula a partir de una matriz de correlación múltiple con correlaciones de difícil interpretación en términos de pocos y más importantes factores, resultando éstos relativamente independientes. En general el método del análisis factorial se resume en cuatro etapas:

¹ Ponce RER, Dickinson BMA, Rodríguez LMA, Kahan E, Mora SJ, Mazón RJJ. Aplicación del Método Análisis Factorial en la Planificación de Servicios Locales de Salud. Documento enviado para su publicación a la Revista de Salud Pública de México, Septiembre 1994

² Existe cierto límite superior dado lo complejo que resulta interpretar análisis simultáneo de diez o más variables, sin embargo, el límite se define por cuestiones prácticas y no por el propio método del análisis factorial.

- Obtención de la matriz de correlación múltiple.
- Cálculo de los factores iniciales³ (extracción de los componentes principales, análisis de la parsimonia e independencia entre factores y porcentaje de varianza⁴ explicada por el modelo).
- Rotación de factores (varimax, equimax y oblicua)
- Cálculo de factores finales y peso absoluto de las variables en cada factor calculado.

Además de lo anterior, los factores se pueden conceptualizar y priori o posteriori ("prueba de hipótesis multivariada") y sus principales aplicaciones se pueden clasificar para: usos exploratorios, confirmatorios y constructivos. Los primeros exploran y detectan patrones de las variables con una orientación hacia el descubrimiento de nuevos conceptos y posibles relaciones de datos; los segundos, prueban hipótesis sobre la estructura de las variables en términos del número esperado de factores significativos y cargas de los factores; los últimos, obtienen mediciones sobre la construcción de índices para usarse como nuevas variables en posteriores análisis.

En este contexto surgen las siguientes preguntas: ¿Cómo emitir un juicio de la calidad de un trabajo de investigación en el cual no se sabe si una técnica de análisis multivariado es la que se debe usar según los objetivos y diseño utilizados?, si sabemos dar respuesta a esta pregunta ¿existen técnicas alternativas?, ¿influye el tamaño de la muestra y el número de observaciones por cada variable para poder usar con seguridad determinada técnica multivariada?, ¿en qué escala de medición y qué distribución deben tener los datos obtenidos?.

El propósito de este ensayo es el de motivar la reflexión y evidenciar que el uso de técnicas de análisis multivariado cada vez es más frecuente y que este hecho debe ser considerado por los Comités Editoriales de las Revistas que evalúan trabajos sometidos para publicación en las mismas; lo anterior deberá repercutir en los juicios evaluatorios de los trabajos de investigación y de esta forma fortalecer la evaluación de los mismos con una base más objetiva.

Aunque la estadística multivariada ya es accesible a toda la comunidad de investigadores en nuestro medio, pocos son los que la usan y aún menos los que eligen y utilizan apropiadamente determinada técnica describiendo en sus publicaciones información detallada y precisa de la misma; además, los autores deben proporcionar información mínima suficiente para que los revisores y evaluadores, así como el lector versado en el tema, puedan emitir un juicio de la calidad del trabajo apoyado con los elementos necesarios para hacerlo.

Por otra parte, el uso de las técnicas multivariadas no es exclusivo para la investigación sociomédica, su aplicación en las áreas clínica y biomédica están más al alcance de lo que muchas personas piensan; la razón fundamental para sostener este concepto es que en estas áreas de investigación frecuentemente se estudian variables de tipo cuantitativo continuo, lo cual es "idóneo" para aplicar las TEAM.

En la investigación experimental de tipo biomédico es frecuente que se controlen con mayor seguridad variables que pueden intervenir o interferir en la respuesta a determinado estímulo, esto permite diferenciar con claridad la participación individual o simultánea de dos, tres o más variables independientes sobre una o más dependientes; las TEAM pueden detectar cuál de todas las variables posee mayor trascendencia dentro del conjunto total.

En la investigación aplicada (clínica), que es un campo muy amplio de investigación en nuestro medio, existe una gran variedad de líneas de investigación en las cuales la descripción de enfermedades, comparación de terapéuticas, evaluación de eficacia y tolerancia de tratamientos, etcétera, proporcionan un terreno accesible y muy propio para aplicar técnicas de análisis multivariado.

Para finalizar este ensayo, proporcionamos referencias adicionales para quien desee iniciar o reforzar sus conocimientos acerca de las técnicas de análisis multivariado¹⁴⁻²¹. Esperamos colaborar al desarrollo de elementos de juicio para evaluar un trabajo sometido a publicación de tal manera que, los lectores de las revistas médicas realicen una lectura más crítica y selectiva.

³ Generalmente deben calcularse dos factores para dar mayor interpretación a los resultados.

⁴ Se considera una buena solución a un modelo que explique más del 50% de varianza con dos factores únicamente.

Referencias

1. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. *Ann Intern Med* 1988;108:258-65.
2. International Steering Committee. Uniform requirements for manuscripts submitted to biomedical journals. *Ann Intern Med* 1979;90:95-9.
3. International Committee of Medical Journals Editors. Uniform requirements for manuscripts submitted to biomedical Journals. *Ann Intern Med* 1982;96:766-71.
4. International Committee of Medical Journals Editors. Uniform requirements for manuscripts submitted to biomedical Journals. *Br Med J* 1991;302:338-41.
5. Comité Internacional de Editores de Revistas Médicas. Requisitos uniformes para preparar los manuscritos que se proponen para publicación en revistas biomédicas. *Bol Of Sanit Panam* 1989;107(5):422-37.
6. Comité Internacional de Editores de Revistas Médicas. Requisitos uniformes para preparar los manuscritos enviados a revistas biomédicas. *Salud Pública Mex* 1992;34(1):94-102.
7. Bailar JC, Mosteller F. La información estadística que deben proporcionar los artículos publicados en revistas médicas. *Salud Publica Méx* 1992;34(1):103-15.
8. Eysenck S, Lara-Cantú MA. Estudio transcultural de la personalidad en niños mexicanos e ingleses. *Salud Pública Mex* 1992;34(1):50-7.
9. Romero GXR, Perham-Zellmer KA, Vázquez CRII, Díaz GA, García MFE, Gómez SRA, Montijo QRE, Aguirre GJS, Izquierdo AMG, Hernández TI, Zapata O. Regionalización en salud: Un instrumento para la planeación jurisdiccional. *Salud Pública Mex* 1992;34(5):506-17.
10. Lara MA, Acevedo M, López EK, Fernández M. La salud emocional y las tensiones asociadas con los papeles de género en las madres que trabajan y en las que no trabajan. *Salud Mental* 1993;16(2):13-21.
11. Laniado-Laborín R, Molgaard CA, Elder JP. Efectividad de un programa de prevención de tabaquismo en escolares mexicanos. *Salud Pública Mex* 1993;35(4):403-8.
12. Miranda-Ocampo R, Salvatierra-Izaba B, Vivanco-Cedeño B, Alvarez-Lucas C, Lezana-Fernández MA. Inequidad de los servicios de salud a población abierta en México. *Salud Pública Mex* 1993;35(6):576-84.
13. Jiménez C, Romieu I, Palazuelos E, Muñoz I, Cortés M, rivero A, Catalán J. Factores de exposición ambiental y concentraciones de plomo en sangre en niños de la ciudad de México, *Salud Pública Mex* 1993;35(6):599-606.
14. Nie Norman H, Hadlai HC, Jenkins J G, Stewbrenner K and Bent DII. *Statistical package for the social sciences*. 2a. edición; New York; McGraw Hill, 1975.
15. Kleinbaum DG and Kupper LI. *Applied regression analysis and other multivariable methods*. Wadworth Publishing Company; California, 1978.
16. Williams F. *Razonamiento estadístico*, 2a. edición; Nueva Editorial Interamericana; México, 1982.
17. Stephen JG. *The Mismeasure of man*. WW Norton Company. New York, 1981.
18. Bailar III JC, Mosteler F. *Medical Uses of Statistics*. NEJM Books, 2nd. Edition. Boston-Massachusetts, USA, 1992.
19. Namakforoosh MN. *Metodología de la investigación*, Limusa, México 1994.
20. Riegelman RK, Hirsch RP. *Cómo estudiar un estudio y probar una prueba: Lectura crítica de la literatura médica*. Parte XII, Serie Comunicación Biomédica. *Bol Of Sanit. Panm* 1992;112(6):521-38.
21. Dawson-Saunders B, Trapp RG. *Basic and Clinical Biostatistics*. Appleton and Lange Medical Book, Norwalk, Connecticut, USA, 1994.