

Análisis de Secuencias Sociales: método novedoso para la medición de desigualdades sociales en salud

ValdésSantiago,Damian¹
Pría Barros, María del Carmen²
Álvarez Lauzarique, María Esther³
Bayarre Veá, Héctor Demetrio⁴

¹Universidad de La Habana/Departamento de Matemática Aplicada, Facultad de Matemática y Computación, La Habana, Cuba, dvs89cs@matcom.uh.cu

²Escuela Nacional de Salud Pública/Departamento de Bioestadística, La Habana, Cuba, mcpria@infomed.sld.cu

²Escuela Nacional de Salud Pública/Departamento de Bioestadística, La Habana, Cuba, esther.alvarez@infomed.sld.cu

²Escuela Nacional de Salud Pública/Departamento de Bioestadística, La Habana, Cuba, hbayarre@infomed.sld.cu

Resumen: Una desigualdad social en salud es una diferencia o variación en el perfil o en los logros de salud de individuos o de población con respecto a grupos sociales, definidos por una variable o un indicador socioeconómico. Para la medición de las desigualdades en salud se han utilizado diversos métodos como los métodos basados en categorías ordenadas, la modelación multinivel y el escalamiento multidimensional. El análisis de secuencias sociales (ASS) es una técnica cuantitativa que puede trabajar con datos de naturaleza cuali-cuantitativa. Este método representa las trayectorias sociales y de salud de las personas mediante secuencias, realizando un agrupamiento por similitud de las mismas para detectar patrones en dichas trayectorias. Este análisis puede realizarse para toda la muestra o por estratos según variables de salud y socioeconómicas, y puede combinarse con otras técnicas como la regresión. Además, brinda formas de visualización de las secuencias, cálculo de medidas de variabilidad de las mismas y grafica las secuencias representativas. En esta ponencia se propone el uso del ASS y su combinación con los métodos tradicionales para la medición de desigualdades en salud. Se destacan las características que hacen que este método sea adecuado para este propósito, se realizan recomendaciones para su uso y se presenta software disponible para su aplicación. Ante los desafíos metodológicos de esta técnica se presentan alternativas para la selección del número de grupos en el agrupamiento, algoritmo de agrupamiento a utilizar y la detección de puntos de inflexión.

Palabras clave: desigualdades sociales en salud, análisis de secuencias sociales, agrupamiento, punto de inflexión, TraMineR.

I. INTRODUCCIÓN

Una desigualdad en salud es una diferencia o variación en el perfil o en los logros de salud de individuos o de una población. El término desigualdad social en salud se refiere a una desigualdad en salud con respecto a grupos sociales, definidos por una variable o un indicador socioeconómico (1). Una inequidad en salud es una desigualdad en salud sistemática (no se distribuye al azar), producida socialmente (modificable) e injusta (se mantiene como producto de órdenes sociales injustos) (2).

En el año 2001, se elaboró un proyecto inicial para monitorear la equidad en salud en Cuba (3). En la Escuela Nacional de Salud Pública se han realizado diversas investigaciones relacionadas con la equidad en salud: reflexiones metodológicas y epistemológicas(4), estudios de género (5), condiciones de vida (6), familia y violencia en la mujer. A estas investigaciones se suman las ejecutadas en otras instituciones (7,8) y las elaboradas por Bacallao y colaboradores (9).

Un indicador para medir desigualdades sociales en salud debe presentar las siguientes características (10): (i) reflejar la dimensión socioeconómica de las desigualdades en el campo de la salud, (ii) incorporar la información correspondiente a todos los grupos de la población definidos por el indicador, y (iii) ser sensible a los cambios en la distribución y al tamaño de la población a lo largo de la escala socioeconómica.

Las investigaciones consultadas verifican que en Cuba existen bajos niveles de covariación de las desigualdades socioeconómicas con la situación de salud. Como alternativa para solucionar este problema se proponen la medición multidimensional de la desigualdad (11), el descubrimiento de variables latentes mediante análisis de conglomerados basado en modelos y en regresión (12), y métodos basados en datos multivariados (9).

Por otra parte, la Epidemiología Social (13) propone diversos enfoques para el estudio de las desigualdades en salud(4,14). La teoría ecosocial de Krieger reconoce que la distribución de los problemas de salud no se explica solamente sobre la base de características individuales sino también por las características de la población y para esto recurre al análisis multinivel(15).

La teoría ecosocial se basa también en la perspectiva del curso de la vida(PCV) (16) para analizar las trayectorias del desarrollo biológico y social de las personas a lo largo del tiempo, de acuerdo al período histórico, el contexto social, económico, político, tecnológico y ecológico que les tocó vivir (13). En la PCV se emplean métodos cualitativos y cuantitativos (16). Entre las técnicas cuantitativas que pueden utilizarse se encuentra el Análisis de Secuencias Sociales(ASS) (17,18).

Hougham y colaboradores utilizaron ASS para estudiar la neumonía extrahospitalaria en hospitales de Estados Unidos, obteniendo que los elementos del grupo de personas que murieron en el hospital tuvieron mayor edad, pasaron más tiempo ingresados y gastaron más en la atención médica que el resto de los pacientes (19). Le Meur y colaboradores utilizaron ASS para estudiar trayectorias en el cuidado institucional de mujeres embarazadas (20). Roux y colaboradores utilizaron esta técnica para describir los cuidados de personas con esclerosis múltiple en Francia, identificando diferencias de los cuidados en relación a la edad de los pacientes (21). Ganjour y colaboradores estudiaron la relación existente entre el género y las trayectorias ocupacionales individuales y los beneficios del retiro laboral (22). En Cuba no está reportado el uso del ASS.

En este artículo se propone el ASS para medir la desigualdad social en salud en Cuba, se explican brevemente los fundamentos del método. Posteriormente, se justifican sus posibilidades para medir la desigualdad, se brindan recomendaciones para su uso, se proponen alternativas a los desafíos metodológicos de la técnica y se presenta software disponible para su aplicación.

II. ANÁLISIS DE SECUENCIAS SOCIALES

La PCV en salud proviene de la sociología y considera la influencia en la salud del contexto histórico, político, socioeconómico y familiar, además de los rasgos biológicos(23). La PCV plantea cinco principios: (i) el desarrollo humano es un proceso que ocurre a lo largo de la vida, multidimensional y multidireccional, (ii) el curso de vida de los individuos está contextualizado en un tiempo histórico y en un espacio, (iii) los individuos construyen su propio curso de vida a través de opciones y acciones, en relación con las oportunidades y las obligaciones impuestas por el contexto, (iv) se vive de manera interdependiente; por eso la trayectoria de un individuo impacta en la de los demás y es influenciada por la de su entorno, y (v) el impacto de las transiciones y de los eventos (individuales o socio-históricos) varía según el ritmo de la vida de una persona, conforme al momento en que ocurran(24).

Entre las técnicas cuantitativas para estudiar en la PCV se encuentran la regresión múltiple, modelos multinivel, modelación mediante ecuaciones estructurales y ASS. Esta técnica constituye una extrapolación de las técnicas de alineación empleadas en el estudio del ADN (17).

El objetivo fundamental del ASS es detectar patrones en las secuencias de eventos, estados y cualquier otro elemento social que pueda observarse para un conjunto determinado de actores. Esto implica definir qué se considera un patrón significativo y cómo detectarlo. Un ASS permite conocer si los cursos de vida obedecen alguna norma social, qué tipo de desenlaces se observan en esos “estándares”, por qué algunas personas tienen mayor riesgo de seguir una trayectoria “caótica”, cómo se relaciona la trayectoria de vida con el sexo, el origen social y otros factores culturales, cómo se relaciona un desenlace como el estado de salud o un ingreso hospitalario con una trayectoria, entre otras.

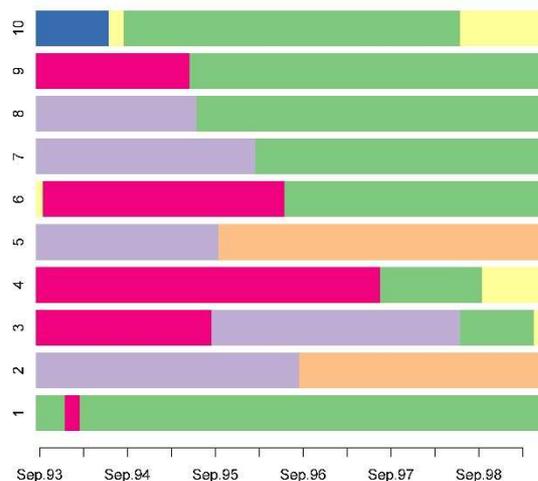


Figura 1: Diez secuencias sociales representadas. Cada barra horizontal visualiza la trayectoria de un individuo respecto a una variable longitudinal. Cada color representa un estado diferente dentro de la trayectoria. En el eje horizontal se tiene los momentos temporales en que fue medida la variable.

Las secuencias sociales son regularidades observadas de forma empírica donde el orden de sus elementos depende del tiempo (18). Una secuencia social representa la trayectoria de una persona (e.g. educación) y describe el decursar en el tiempo del individuo respecto a una faceta de su vida (Fig. 1). En cada trayectoria y con cierta periodicidad, se pueden alcanzar una serie de estados o alfabeto (e.g.

primaria, secundaria, etc.) codificado mediante caracteres simples (e.g. P, S, etc.). El cambio entre estados consecutivos se denomina transición.

Sobre estas secuencias se aplican algoritmos para obtener una matriz de similitud entre todos los pares de secuencias en la población. Luego, se agrupan las secuencias y se realiza la interpretación de agrupamiento por parte del especialista. Este proceder se conoce como *CoreProgram* y fue propuesto por Abbott en 1986 (17). En este algoritmo se utiliza el método *OptimalMatching* para calcular la similitud entre las secuencias y el agrupamiento jerárquico de Ward.

III. RECOMENDACIONES PARA USAR ASS EN LA MEDICIÓN DE DESIGUALDADES SOCIALES EN SALUD

Las posibilidades de esta técnica la hacen adecuada para la detección de desigualdades sociales en salud y cumple con los criterios de Wagstaff ya citados. El ASS refleja la dimensión socioeconómica pues permite la inclusión en el análisis de datos cuantitativos y cualitativos asociados a dicha dimensión. También permite incorporar información de todos los grupos (o de una parte) de la muestra considerada. Además, este método es sensible a los cambios en la distribución y al tamaño de la población a lo largo de la escala socioeconómica. Esta sensibilidad se manifiesta en el resultado del agrupamiento de las secuencias, el estudio de la dinámica de la población mediante gráficos de distribución y entropía, la posibilidad de correlacionar las variables y la detección de secuencias patrón (19,25). A continuación se expondrán algunas recomendaciones para el uso del ASS en la medición de las desigualdades.

Para la realización del agrupamiento se necesita un criterio de similitud entre las secuencias correspondientes a cada individuo. Este proceso implica un cómputo de la similitud o distancia entre cada par de secuencias, que depende de las longitudes de las secuencias involucradas en el cálculo, obteniéndose una matriz de similitudes entre las secuencias. Mientras mayor sea el número de elementos en las secuencias será más costoso, desde el punto de vista computacional, el cálculo de la similitud. Mientras mayor sea la cantidad de secuencias en la muestra, este costo también aumentará. Las distancias reportadas en la literatura y sus ventajas pueden consultarse en (18,26).

A partir de la matriz de similitudes obtenida se puede ejecutar un agrupamiento jerárquico partiendo de los individuos. En dicho agrupamiento los conglomerados están estructurados según diferentes niveles de similitud. Por ello se necesita un criterio para encontrar la menor cantidad de grupos que más se diferencien. Uno de los criterios que pueden utilizarse para ello se denomina *AverageSilhouetteWidth* (27). Si esta medida de calidad del agrupamiento está entre 0,71 y 1,00 se considera que la diferenciación es excelente y si es menor que 0,25 la separación es muy mala (28).

Existen varias formas de realizar el agrupamiento entre secuencias. Dlouhy y Biemann (29) plantean una comparación entre varios métodos de agrupamiento y concluyen que los mejores resultados se logran con el método de Ward.

De la matriz de similitudes puede obtenerse información sobre la representatividad de las secuencias: cuál es la secuencia promedio o central (“norma”), medir la diferencia (*discrepancy*) entre las secuencias y obtener secuencias representativas para toda la muestra o por grupos poblacionales. Además, se puede probar y evaluar la relación entre una o varias covariables y las secuencias mediante análisis de regresión y árboles de regresión (30).

El ASS permite visualizar de forma compacta la dinámica de la población y sus grupos mediante los gráficos como el *sequenceindexplot*, conocer las secuencias que más se repiten, las transiciones más

comunes (30), medidas de la variabilidad de las transiciones de las secuencias (31), correlacionar los grupos con variables socioeconómicas y de salud (30), entre otras posibilidades (32). También puede realizarse la minería de eventos (30) y las redes sociales como métodos más sofisticados de análisis dentro del ASS.

El ASS presenta desafíos metodológicos relativos. Este método requiere datos cualitativos longitudinales para su utilización aunque no necesariamente deben observarse los individuos todo el tiempo(18). La recogida de los datos puede hacerse de forma continua mediante cuestionarios específicos, escalas, encuestas online, datos capturados en tiempo real por dispositivos móviles, datos obtenidos de las redes sociales como Facebook o Twitter y registros médicos. En dependencia de la forma de recolección puede ser necesario un preprocesamiento de los datos para que puedan ser utilizados por el ASS. Una cuestión importante sobre la recolección es que debe hacerse con la misma periodicidad para todos los individuos, aunque la periodicidad no tiene que ser constante su elección depende del problema de salud estudiado.

En esta técnica, un agrupamiento incorrecto depende principalmente del tamaño de la secuencias. Se aconseja usar secuencias con al menos 25 elementos. Las secuencias de menor longitud deben tener al menos el 70% de la longitud de la secuencia más grande. El tamaño de la muestra no afecta sustancialmente la calidad del agrupamiento. Una secuencia con menos del 30% de elementos faltantes no empeora significativamente la calidad del agrupamiento (29).

El ASS presenta debates epistemológicos asociados a cómo identificar mediante esta técnica los fenómenos de la PCV como el concepto de punto de inflexión. Un punto de inflexión es un momento en el que se produce un cambio decisivo en una situación, especialmente uno con resultados beneficiosos. Por ejemplo: una enfermedad propia o de un familiar, pérdida de un pariente en la infancia, una viudez precoz, una crisis existencial, un accidente, crisis económicas, guerras, dictaduras y catástrofes naturales (33). Como aporte de este artículo se proponen los siguientes criterios para identificar puntos de inflexión: transición de mayor costo en una secuencia, transición que provoca cambios inmediatos en el resto de las secuencias y en un conjunto de las mismas, y transición que aumenta la “complejidad” del resto de las secuencias a partir del momento de ocurrencia.

Respecto al software disponible para ejecutar un ASS están reportados en la literatura: programa *Optimize*(34) al que no se le está dando soporte, TDA(35) que es software libre, el paquete SQ del software Stata(36) que está disponible libremente para usuarios con licencia de Stata, CHESA software libre creado por Elzinga(37), SADI (38) y TraMineR(39). TraMineR permite la manipulación, descripción e interpretación de secuencias, y en general la minería de datos secuenciales para las ciencias sociales. Esta librería acepta diferentes representaciones de secuencias y tiene herramientas que permiten la conversión entre formatos. TraMineR presenta varias funciones para computar distancias entre secuencias, análisis de disimilitudes, conocer las subsecuencias más frecuentes e identificar las más discriminantes (40). Esta librería está incluida en el lenguaje estadístico R por lo que es de libre uso y puede usarse en varios sistemas operativos. La librería cuenta con una ayuda local y con foros de discusión online. Según Cornwell, todavía no se cuenta con un software “amigable” que permita a salubristas y sociólogos el uso fácil del ASS (18).

IV. CONCLUSIONES

La desigualdad social en salud es un tema importante y ha sido tratado por investigadores cubanos, sin embargo, el tema de la medición de la desigualdad ha sido menos estudiado. En este artículo se propuso el ASS para la medición de desigualdades en salud, planteando su conveniencia para esto y realizando recomendaciones para su uso como la selección del método de agrupamiento, las formas de visualizar la información secuencial, la determinación de puntos de inflexión y software disponible para su aplicación. Se recomienda combinar las técnicas “clásicas” para medir la desigualdad en el marco secuencial, tomando de ellas ventajas y minimizando las desventajas según el problema.

REFERENCIAS

1. Kawachi I, Subramanian S V, Almeida-Filho N. A glossary for health inequalities. *J Epidemiol Community Heal.* 2002;56:647–652.
2. Whitehead M, Dahlgren G. Concepts and Principles for Tackling Social Inequities in Health: Levelling up Part 1. 2006;(2).
3. Ramírez A, López C. A propósito de un sistema de monitoreo de la equidad en salud en Cuba. *Rev Cuba Salud Pública.* 2005;31(2).
4. Algunas reflexiones acerca de la noción de determinantes, determinantes sociales y desigualdades en salud [Internet]. [cited 2015 Dec 14]. Available from: <http://bvs.sld.cu/revistas/inf/n1512/inf04212.htm>
5. Castañeda IE. Construcción de indicadores sintéticos para medir diferencias de género en el contexto social cubano. Tesis en opción al grado de Doctora en Ciencias de la Salud, Escuela Nacional de Salud Pública, Universidad de Ciencias Médicas de La Habana; 2005.
6. Pría M del C. Diseño de una metodología para el Análisis de la Situación de Salud municipal según condiciones de vida. *Rev Cuba Med Gen Integr.* 2006;22(4).
7. Íñiguez L. Territorio y contextos en la salud de la población. *Rev Cuba Salud Pública.* 2008;34(1).
8. Espina M, Echevarría D. Cuba: los correlatos socioculturales del cambio económico. Editorial de Ciencias Sociales; 2015.
9. Bacallao J. Ensayo crítico acerca de la medición de las desigualdades sociales en salud. Editorial de Ciencias Médicas; 2015.
10. Wagstaff A, Paci P, Doorslaer E Van. On the measurement of inequalities in health. *Soc Sci Med.* 1991;33(5):545–57.
11. Yalonetzky G. A dissimilarity index of multidimensional inequality. *J Econ Inequal.* 2012;(10):343–73.
12. Qian G, Wu Y, Shao Q. A procedure for estimating the number of clusters in logistic regression clustering. *J Classif.* 2009;(26):183–99.
13. Krieger N. Glosario de epidemiología social. *Rev Panam Salud Pública.* 2002;11(5/6):480–90.
14. Ramis Andalia RM. Una vez más, de la vieja a la nueva epidemiología. *Rev Cuba Salud Pública.* 2011;37(4).
15. Kondo N, Sembajwe G, Kawachi I, Dam RM Van, Subramanian S V, Yamagata Z. Income inequality, mortality, and self rated health: meta-analysis of multilevel studies. *BMJ.* 2009;339(b4471):1178–82.

16. Shanahan MJ, Mortimer JT, Kirkpatrick M. Handbook of the Life Course Volume II. Vol. II. Springer International Publishing Switzerland; 2016.
17. Abbott A, Forrest J. Optimal Matching Methods for Historical Sequences. *J Interdiscip Hist.* 1986;16(3):471–94.
18. Cornwell B. Social Sequence Analysis: Methods and Applications. Cambridge University Press; 2015.
19. Hougham GW, Ham SA, Ruhnke GW, Schulwolf E, Auerbach AD, Schnipper JL, et al. Sequence Patterns in the Resolution of Clinical Instabilities in Community-Acquired Pneumonia and Association with Outcomes. *J Gen Intern Med.* 2013;29(4):563–71.
20. Le Meur N, Gao F, Bayat S. Mining care trajectories using health administrative information systems: the use of state sequence analysis to assess disparities in prenatal care consumption. *BMC Health Serv Res.* 2015;15(200):1–10.
21. Roux J, Le Meur N, Grimaud O, Leray E. Care pathways of patients affected with multiple sclerosis in France from 2007 to 2013 using administrative databases and state sequence analysis. In: Ritschard G, Studer M, editors. International Conference on Sequence Analysis and Related Methods (LaCOSA II). 2016. p. 191–206.
22. Ganjour O, Gauthier J-A, Goff J-M Le. Gender inequality regarding retirement benefits in Switzerland. In: Ritschard G, Studer M, editors. International Conference on Sequence Analysis and Related Methods (LaCOSA II). 2016. p. 693–716.
23. Elder GHJ, Kirkpatrick Johnson M, Crosnoe R. Handbook of the Life Course Theory. Handbook of the Life Course. 2003.
24. Mortimer JT, Shanahan MJ. Handbook of the Life Course. Kluwer Academic Publishers; 2002.
25. Hougham GW, Ham SA, Ruhnke GW, Schulwolf E, Auerbach AD, Schnipper JL, et al. Online Supplement to accompany: Sequence Patterns in the Resolution of Clinical Instabilities in Community - Acquired Pneumonia and Association with Outcomes. *J Gen Intern Med.* 2013;29(4):1–9.
26. Studer M. What matters in differences between life trajectories□: a comparative review of sequence dissimilarity measures. *J R Stat Soc.* 2016;(December).
27. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math.* 1987;20:53–65.
28. Gan G, Ma C, Wu J. Data Clustering: Theory, Algorithms, and Applications. American Statistical Association and the Society for Industrial and Applied Mathematics; 2007.
29. Dlouhy K, Biemann T. Optimal matching analysis in career research: A review and some best-practice recommendations. *J Vocat Behav.* 2015;90:163–73.
30. Ritschard G, Gabadinho A, Nicolas M, Studer M. Mining event histories: A social science perspective. *Int J Data Mining, Model Manag.* 2008;1(1):68–90.
31. Winkle Z Van, Fasang A. Complexity in Employment Life Courses in Europe in the Twentieth Century — Large Cross-National Differences but Little Change across Birth Cohorts. *Soc Forces.* 2017;1–29.
32. Helske S. Statistical analysis of life sequence data. University of Jyväskylä; 2016.
33. Abbott A. Time Matters: On Theory and Method. The University of Chicago; 2001.
34. Abbott A. Optimize: Optimal Matching. 1997.
35. Rohwer G, Potter U. TDA: Transition Data Analysis. 2002.

36. Brzinsky-fay C, Kohler U, Luniak M. Sequence analysis with Stata. *Stata J.* 2006;6(4):435–60.
37. Elzinga CH. CHESA 2.1 User Manual. Vrije Universiteit, Amsterdam; 2007.
38. Halpin B. SADI: Sequence Analysis Tools for Stata. 2014. Report No.: WP2014-03.
39. Gabadinho A, Ritschard G, Studer M. Analyzing and Visualizing State Sequences in R with TraMineR. *J Stat Softw.* 2011;40(4).
40. Gabadinho A, Ritschard G, Studer M, Müller NS. Mining sequence data in R with the TraMineR package: A user's guide. Vol. 1. University of Geneva; 2010.