

## COMPARACIÓN DE TASAS DE EVENTOS EPIDEMIOLÓGICOS ENTRE DOS POBLACIONES A TRAVÉS DEL ANÁLISIS EDAD-PERÍODO-COHORTE

PRUNELLO, MARCOS MIGUEL<sup>1</sup>

HACHUEL, LETICIA SUSANA<sup>2</sup>

ARNESI, NORA ELBA<sup>3</sup>

<sup>1</sup> IITAE. Fac. de Ciencias Económicas y Estadística. UNR, [marcosprunello@gmail.com](mailto:marcosprunello@gmail.com)

<sup>2</sup> IITAE. Fac. de Ciencias Económicas y Estadística. UNR, [lhachuel@gmail.com](mailto:lhachuel@gmail.com)

<sup>3</sup> IITAE. Fac. de Ciencias Económicas y Estadística. UNR, [narnesi@gmail.com](mailto:narnesi@gmail.com)

### RESUMEN

En toda población es de gran importancia el estudio de la ocurrencia de ciertos eventos epidemiológicos, como la mortalidad por distintas causas, la cual se ve influenciada por la edad de los individuos, la cohorte a la que pertenecen y el período en el cual se registra el evento. La tarea de desentrañar la influencia compleja e interactiva de estas variables estrechamente relacionadas se denomina Análisis Edad-Período-Cohorte (EPC). Generalmente, el objetivo es estimar en forma separada los efectos asociados a estos tres factores y adjudicar la variabilidad del fenómeno estudiado a cada uno de ellos. Los modelos estadísticos EPC se enfrentan con el denominado *problema de la identificación* debido a que las variables mantienen una relación lineal exacta que impide realizar un ajuste tradicional y no se halla en la literatura estadística un consenso general para su solución. Recientemente se ha desarrollado una metodología útil para estudios en los que el interés es comparar las tasas de un evento entre dos poblaciones. La misma contrasta los parámetros de los modelos EPC ajustados en cada grupo utilizando el enfoque de las funciones estimables como alternativa al problema de la identificación. En este contexto, es posible definir tests de hipótesis estadísticas para determinar si las razones de tasas entre las poblaciones son absolutamente proporcionales o proporcionales dentro de alguno de los tres factores: edad, período o cohorte. Para este trabajo se analizaron las tasas de mortalidad por dieciocho causas de muerte en Argentina, comparándolas entre hombres y mujeres, de 25 a 79 años de edad, entre los años 1980 y 2009. En el 78% de los casos los modelos EPC proporcionaron un ajuste adecuado permitiendo realizar los tests de proporcionalidad y evidenciando que esta

X CONGRESO LATINOAMERICANO DE SOCIEDADES DE ESTADÍSTICA  
CÓRDOBA, ARGENTINA. 16 A 19 DE OCTUBRE 2012

metodología resulta un aporte beneficioso como instrumento para organizar y clarificar un análisis comparativo EPC.

***PALABRAS CLAVE:*** modelos edad-período-cohorte; funciones estimables; proporcionalidad

## 1. INTRODUCCIÓN

En toda población es de gran importancia el estudio de los determinantes y de la evolución en el tiempo de ciertos eventos epidemiológicos como, por ejemplo, la mortalidad por distintas causas. Más aún, en muchas ocasiones las investigaciones se centran en la comparación de la ocurrencia de dichos eventos en dos poblaciones distintas, definidas por características geográficas, sociales, económicas, de género, etc., con el fin de evaluar la influencia de la *edad* de los individuos, la *cohorte* de nacimiento o generación a la que pertenecen y el momento o *período* en el cual se registra el evento. La tarea de desentrañar la influencia compleja e interactiva de estas variables estrechamente relacionadas -*edad, período* y *cohorte*- sobre un proceso epidemiológico o de cambio social, suele denominarse en la literatura estadística “Análisis Edad-Período-Cohorte”, o más sintéticamente, “Análisis EPC”. Generalmente, el objetivo que se persigue es evaluar por separado el impacto de cada una de estas variables en la evolución de las tasas del evento considerado.

Tanto los efectos asociados a período como a cohorte pretenden explicar los cambios en dichas tasas vinculados con el transcurso del tiempo, ambos importantes en el intento de establecer tendencias en las tasas de incidencia de una enfermedad o de mortalidad. Los efectos asociados a período pueden estar ligados a descubrimientos científicos o a cambios de costumbres, es decir, a situaciones que afectan a la ocurrencia del evento, impactando sobre la población entera (por ejemplo, la aparición de un nuevo procedimiento terapéutico en un momento dado puede modificar el pronóstico de una enfermedad simultáneamente en toda la población). En cambio, los efectos asociados a cohorte reflejan el impacto de hábitos de largo plazo o exposiciones de mucho tiempo, haciendo que diferentes generaciones estén expuestas a distintos niveles de riesgo (por ejemplo, generaciones expuestas a diferentes grados de contaminación). En cuanto a la edad, es claro que constituye un factor de riesgo trascendente para muchas variables relacionadas a la salud (por ejemplo, la aparición de enfermedades crónicas suele incrementarse al aumentar la edad).

Tradicionalmente, el análisis EPC se ha abordado a través de herramientas gráficas y descriptivas (Robertson, 1998) que son útiles en una primera instancia exploratoria pero que resultan insuficientes para dar respuestas formales a muchos interrogantes que pueden surgir. Por este motivo, paralelamente se han propuestos modelos estadísticos, denominados modelos EPC (Clayton y Schiffers, 1987; Holford, 1983,1991), basados en la teoría de los modelos lineales generalizados (Agresti, 2002), con los cuales se intenta estimar en forma separada los

efectos asociados a los tres factores con el fin de adjudicar a cada uno de ellos la variabilidad del fenómeno bajo estudio.

Sin embargo, esta modelización es de uso poco frecuente en la práctica principalmente debido a la falta de una completa comprensión del *problema de la identificación*: edad, período y cohorte no son variables independientes, sino que mantienen una relación lineal exacta: la diferencia entre el *período* y la *edad* es igual a la *cohorte*. Este hecho impide que el ajuste tradicional de modelos estadísticos obtenga una estimación para el efecto de cada nivel de los tres factores y no existe un consenso generalizado en la literatura para su tratamiento.

Recientemente ha aparecido un enfoque que utiliza el ajuste de modelos EPC en dos poblaciones con el fin de comparar entre ellas el comportamiento de las tasas de ocurrencia del evento de interés. Este método soluciona el problema de la identificación utilizando una reparametrización de los modelos en términos de funciones estimables y evalúa, a través de pruebas de hipótesis estadísticas sobre los parámetros, la existencia de algún tipo de proporcionalidad entre dichas tasas, la cual puede manifestarse de forma absoluta o luego de estratificar por *edad*, *período* o *cohorte* (Rosenberg y Anderson, 2010, 2011).

En este trabajo se presenta la metodología propuesta por este enfoque y los resultados de la aplicación del mismo sobre tasas de mortalidad por distintas causas en Argentina, con el objetivo de demostrar su utilidad y su capacidad como una herramienta formal que enriquece ampliamente el estudio comparativo de tasas de eventos epidemiológicos entre dos poblaciones.

## 2. METODOLOGÍA

A partir de los Registros de Estadísticas Vitales (Dirección de Estadísticas e Información de Salud, Ministerio de Salud de la Nación) y de los Registros Censales (INDEC) se cuenta con información sobre el número de muertes por distintas causas para intervalos quinquenales de edad y de período, y sus respectivas cantidades de personas-año necesarias para la construcción de tasas.

Un Modelo Lineal Generalizado (MLG) adecuado para el estudio de las tasas de mortalidad es:

$$\ln \left( E \left( \frac{y_{ep}}{N_{ep}} \right) \right) = \ln \left( \frac{\mu_{ep}}{N_{ep}} \right) = \lambda + \alpha_e + \pi_p + \gamma_c \quad [1]$$

donde  $y_{ep}$  es el número de muertes en el período calendario  $p$  y el grupo etario  $e$  y  $N_{ep}$  es el correspondiente número de personas-año en riesgo, con  $e = 1, \dots, E$  y  $p = 1, \dots, P$ . Además,  $\lambda$

es la media general y  $\alpha_e$ ,  $\pi_p$  y  $\gamma_c$  son los efectos principales del  $e$ -ésimo grupo de edad, del  $p$ -ésimo período de tiempo y de la  $c$ -ésima cohorte de nacimiento. El subíndice  $c$  queda automáticamente determinado a través de la relación  $c = p - e + E$  ( $c = 1, \dots, C = P + E - 1$ ). El número de casos  $y_{ep}$  se considera como una realización de una variable aleatoria Poisson, con media  $\mu_{ep}$ .

En el ajuste de un MLG se suele emplear como restricción que la suma de los efectos de los niveles de cada factor sea igual a cero o considerar el efecto de un nivel de cada factor igual a cero. Estas restricciones habituales no solucionan el *problema de la identificación* que surge debido a la exacta relación lineal existente entre las variables: *cohorte = período - edad*. Es necesario, entonces, implementar una restricción adicional la cual podría tratarse, por ejemplo, de la igualdad entre los efectos de dos niveles en uno de los factores.

Con las restricciones habituales y la restricción adicional impuesta es posible hallar un único conjunto de estimaciones máximo verosímiles de los parámetros del modelo. Sin embargo, los resultados que se obtienen dependen de esta restricción adicional, por lo que para su elección se requiere un profundo conocimiento del proceso estudiado, ya que si la misma no se corresponde con la realidad, se incurriría en interpretaciones sesgadas.

A lo largo de los años se han desarrollado distintas soluciones al problema de la identificación de manera de evitar la arbitrariedad que en general conlleva dicha imposición. Entre ellas, una de las de mayor popularidad es la propuesta por Holford (1983, 1991), que consiste en limitar el análisis a combinaciones lineales o no lineales de los parámetros que permanecen invariantes con cualquiera de las restricciones adicionales que se imponga, conocidas como *funciones estimables*. Específicamente, este enfoque se centra en la partición de los efectos de las variables en dos componentes: una parte lineal y otra parte de curvatura o desviación de la linealidad, dando lugar al modelo [2]:

$$\ln\left(\frac{\mu_{ep}}{N_{ep}}\right) = \lambda + \alpha_L(e - \bar{e}) + \pi_L(p - \bar{p}) + \gamma_L(c - \bar{c}) + \tilde{\alpha}_e + \tilde{\pi}_p + \tilde{\gamma}_c \quad [2]$$

donde  $\alpha_L$ ,  $\pi_L$  y  $\gamma_L$  son las componentes lineales;  $\tilde{\alpha}_e$ ,  $\tilde{\pi}_p$  y  $\tilde{\gamma}_c$  son las desviaciones de la linealidad o curvaturas; y  $(e - \bar{e})$ ,  $(p - \bar{p})$  y  $(c - \bar{c})$  son índices que representan cada grupo de edad, período y cohorte, respectivamente, con  $\bar{e} = \frac{E+1}{2}$ ,  $\bar{p} = \frac{P+1}{2}$  y  $\bar{c} = \frac{C+1}{2} = \bar{p} - \bar{e} + E$ .

Las curvaturas constituyen funciones estimables, es decir satisfacen la condición de ser invariantes a la parametrización inicial. Por el contrario, las componentes lineales no lo son, pero sí lo son algunas combinaciones lineales de las mismas, en particular las siguientes tres que revisten de un especial interés:

- Tendencia Longitudinal de Edad (TLE):  $\alpha_L + \pi_L$
- Tendencia Transversal de Edad (TTE):  $\alpha_L - \gamma_L$
- Drift Neto (DN):  $\pi_L + \gamma_L$

De acuerdo a estas definiciones el modelo [2] puede reescribirse en términos de las funciones estimables de dos formas diferentes, conocidas como *edad-cohorte* [3] y *edad-período* [4]:

$$\ln\left(\frac{\mu_{ep}}{N_{ep}}\right) = \lambda + (\alpha_L + \pi_L)(e - \bar{e}) + (\pi_L + \gamma_L)(c - \bar{c}) + \tilde{\alpha}_e + \tilde{\pi}_p + \tilde{\gamma}_c \quad [3]$$

$$\ln\left(\frac{\mu_{ep}}{N_{ep}}\right) = \lambda + (\alpha_L - \gamma_L)(e - \bar{e}) + (\pi_L + \gamma_L)(p - \bar{p}) + \tilde{\alpha}_e + \tilde{\pi}_p + \tilde{\gamma}_c \quad [4]$$

Si el objetivo que se persigue es comparar el comportamiento de las tasas de un determinado evento entre dos poblaciones, se ajusta el modelo EPC en cada una de ellas y se evalúa la bondad de dichos ajustes teniendo en cuenta la posible presencia de *sobredispersión*. Luego, es posible buscar evidencia de algún tipo de proporcionalidad entre las tasas de ambos grupos, probando la igualdad de ciertos parámetros de los modelos EPC correspondientes a cada subpoblación mediante tests estadísticos de tipo Wald o de Razón de Verosimilitud. Existen cuatro clases de proporcionalidad posibles:

- **Proporcionalidad absoluta:** se dice que entre ambos conjuntos de tasas existe *proporcionalidad absoluta* si la razón de tasas para un grupo relativo al otro es constante a través de todas las edades, períodos y cohortes. No deben haber diferencias significativas entre los parámetros EPC de ambos grupos, excepto por los interceptos.
- **Proporcionalidad dentro de cohortes:** se presenta si la razón de tasas entre ambos grupos es igual a una constante que depende sólo de la cohorte considerada. Los parámetros que deben ser iguales son la TLE y las desviaciones de edad y período.
- **Proporcionalidad dentro de grupos de edad:** se verifica si la razón de tasas entre ambos grupos es igual a una constante que depende sólo del grupo de edad. Los parámetros que deben ser iguales son el DN y las desviaciones de cohorte y período.
- **Proporcionalidad dentro de períodos:** se evidencia si la razón de tasas entre ambos grupos es igual a una constante que depende sólo de período. Los parámetros que deben ser iguales son la TTE y las desviaciones de edad y cohorte.

Para probar la existencia de alguno de estos cuatro tipos de proporcionalidad se realiza un test de Wald, a través de la estadística chi-cuadrado:

$$X^2 = [\mathbf{L}(\hat{\Psi}^1 - \hat{\Psi}^0)]^T [\mathbf{L}(\widehat{\mathbf{var}}(\hat{\Psi}^1) + \widehat{\mathbf{var}}(\hat{\Psi}^0)) \mathbf{L}^T]^{-1} [\mathbf{L}(\hat{\Psi}^1 - \hat{\Psi}^0)] \quad [5]$$

donde  $\hat{\Psi}^j$  es el vector con las estimaciones de los parámetros del modelo EPC en la población  $j$ ,  $\widehat{\mathbf{var}}(\hat{\Psi}^j)$  es la correspondiente matriz estimada de variancias y covariancias y  $\mathbf{L}$  es una matriz de contraste adecuada para cada test, con  $2(E + P - 2)$  columnas y tantas filas como parámetros se comparen entre ambos grupos ( $j = 0,1$ ). Los grados de libertad corresponden al número de parámetros que deben ser iguales bajo el tipo de proporcionalidad propuesto.

Otra alternativa es realizar un test de Razón de Verosimilitud, el cual compara la bondad de ajuste de dos modelos,  $M_1$  y  $M_0$ . El primero,  $M_1$ , es un modelo más general que contempla que todos los parámetros del modelo EPC son distintos entre ambos grupos poblacionales. Se ajusta sobre el conjunto total de datos a través de las formas [3] o [4], pero incluyendo además un término para el efecto principal *grupo* y las interacciones entre éste y todos los demás parámetros. El segundo,  $M_0$ , es un modelo más simple, análogo al anterior pero no incluye las interacciones entre la indicadora de grupo y aquellos parámetros que deben ser iguales en ambas poblaciones para que se verifique el tipo de proporcionalidad evaluado. La estadística que se emplea es:

$$RV = D_0 - D_1 \quad [6]$$

donde  $D_j$  es la deviance del modelo  $M_j$  ( $j = 0,1$ ). El número de grados de libertad es la diferencia entre el número de parámetros del modelo completo  $M_1$  y el número de parámetros del modelo más simple  $M_0$ .

### 3. RESULTADOS

Para estudiar la utilidad de la metodología descrita se contrastaron las tasas de mortalidad por dieciocho causas de muerte en Argentina entre las poblaciones definidas por género, esto es, para hombres y mujeres, de 25 a 79 años de edad, entre los años 1980 y 2009. En el 78% de los casos los modelos EPC proporcionaron un ajuste adecuado permitiendo llevar a cabo los tests de proporcionalidad. Se encontró algún tipo de proporcionalidad en el 21% de las causas de muerte, con tasas heterogéneas en el 79% restante. Por ejemplo, las tasas de mortalidad por mal de Chagas resultaron ser absolutamente proporcionales, con una razón de tasas estimada hombres versus mujeres de 1.99, lo cual significa que la mortalidad es un 99% mayor para el sexo masculino que para el femenino, en cualquier grupo etario, período de tiempo o cohorte de nacimiento. En el caso de la mortalidad por anemias, se

evidencia proporcionalidad dentro de cohortes, es decir, la razón de tasas sólo depende de la cohorte considerada, observándose que la mortalidad fue superior en hombres durante generaciones pasadas. En cambio, las tasas de mortalidad por leucemia no mostraron ningún tipo de proporcionalidad, por lo que las razones de tasas son significativamente heterogéneas cualquiera sea el grupo de edad, el período de registro o la cohorte de nacimiento. Más allá de que se evidencie o no algún tipo de proporcionalidad, las tasas y razones de tasas ajustadas siempre resultaron útiles, ya que son versiones suavizadas de las tasas crudas permitiendo destacar los aspectos más relevantes. En líneas generales, esta metodología proporciona elementos formales para la comparación de tasas y resulta una útil extensión de los métodos estándares disponibles en la epidemiología descriptiva. Si bien no los reemplaza, pues dichas herramientas tradicionales son populares y proveen resultados apreciables, permite organizar y clarificar el análisis EPC comparativo.

#### 4. REFERENCIAS

- AGRESTI, A. (2002). *An Introduction to Categorical Data Analysis*. Wiley. New York.
- DIRECCIÓN DE ESTADÍSTICAS E INFORMACIÓN DE SALUD (1980-2009). Anuarios de Estadística Vitales, Serie 5. Ministerio de Salud de la Nación. Ciudad Autónoma de Buenos Aires, Argentina.
- INDEC. Estimaciones y Proyecciones de Población. Total del país. 1950-2015. Serie Análisis Demográfico. <http://www.indec.mecon.gov.ar>. Fecha de consulta: 23 de septiembre de 2011.
- ROBERTSON, C. and BOYLE, P. (1998). Age-Period-Cohort analysis of chronic disease rate II: graphical approaches. *Statistics in Medicine*, 17, 1325-1339.
- ROSENBERG, P. and ANDERSON, W. (2010). Proportional hazards models and age-period-cohort analysis of cancer rates. *Statistics in Medicine*, 29, 1228:1238.
- ROSENBERG, P. and ANDERSON, W. (2011). Age-Period-Cohort Models in Cancer Surveillance Research: Ready for Prime Time? *Cancer Epidemiol Biomarkers Prev*, 20, 1263:1268.