

Análisis de independencia entre la causa de muerte y los factores asociados a esta en Costa Rica

Alejandro Brenes Calderón¹, Santiago Fernández Sáenz¹, María Paula Jiménez Torres¹,
Eyeri Méndez Méndez¹

alejandro.brenescalderon@ucr.ac.cr, santiago.fernandezsaenz@ucr.ac.cr,
maria.jimeneztorres@ucr.ac.cr, eyeri.mendez@ucr.ac.cr

RESUMEN

El estudio realizado en Costa Rica se centra en analizar las variables que influyen en las causas de muerte, explorando su dependencia. Utilizando datos de defunciones recopilados entre 2014 y 2021 por el INEC, se construyeron tablas de contingencia entre la causa de muerte (variable principal) y diversas variables demográficas y sociales, tales como sexo, estado civil, ocupación, región de residencia del Ministerio de Salud Pública (MSP), rango etario y nacionalidad. Se aplicó la prueba de independencia χ^2 para cada tabla de contingencia, evaluando la independencia entre la causa de muerte y cada variable. Los resultados mostraron que, en la mayoría de los casos, el p-valor obtenido fue cercano a cero o exactamente cero, lo que confirma una fuerte dependencia entre la causa de muerte y las variables analizadas. Además del análisis de independencia, se utilizó la V de Cramer para cuantificar la fuerza de esta dependencia. Los hallazgos sugieren que factores como el sexo, estado civil, ocupación, región del MSP, rango etario y nacionalidad podrían influir significativamente en las causas de muerte de los fallecimientos ocurridos en territorio nacional. En conclusión, este estudio subraya la importancia de considerar los factores mencionados con anterioridad al diseñar políticas de salud pública, dirigidas a mitigar los riesgos de mortalidad en diferentes áreas de Costa Rica.

PALABRAS CLAVE: prueba de chi cuadrado, factores demográficos, tablas de contingencia, mortalidad.

INTRODUCCIÓN

Cada día se producen fallecimientos de diversas causas a nivel nacional y cada muerte es única, reflejando la particularidad de cada persona por múltiples condiciones. Los gobiernos recopilan las muertes, anotando no solo la causa del fallecimiento, también, variables adicionales de la muerte, como lo puede ser la fecha y el lugar, asimismo, características relacionadas con el fallecido, como podrían ser la edad, el sexo, la ocupación o el estado civil. Con esta abundancia de datos, surge una pregunta: ¿es posible que estas variables tengan algún tipo de relación con las causas de los fallecimientos?

Utilizando los registros de muertes en Costa Rica entre los años 2014 y 2021 (Instituto Nacional de Estadística y Censos, 2023), se propone realizar un análisis de

¹ Estudiante de Ciencias Actuariales de la Universidad de Costa Rica

independencia entre las causas de muerte y los factores asociados a estas en el país.

La importancia de realizar esta investigación radica en su potencial para dar una mejor comprensión de las causas de mortalidad y los factores que influyen en ellas. Este conocimiento es crucial para revelar patrones y tendencias que no son evidentes. Esta investigación busca evidenciar la relación que tiene la causa de muerte con las diversas características del individuo fallecido. Estos hallazgos pueden ser utilizados para desarrollar estrategias de salud pública más eficaces, adaptadas a las necesidades específicas de la población. Además, el análisis de independencia entre las variables y las causas de muerte proporcionará una base sólida para futuros estudios y políticas de salud.

La relevancia política y social de esta investigación es significativa. En un contexto donde los recursos de salud pública son limitados, es fundamental repartirlos de manera eficiente y equitativa. Este estudio puede guiar la distribución de recursos y la implementación de programas de salud, priorizando a aquellos grupos más vulnerables. De modo que el presente documento puede contribuir a reducir las desigualdades en salud, asegurando que todos los habitantes de Costa Rica, independientemente de sus características individuales, tengan una mejor calidad de vida.

Para identificar cuáles variables son influyentes en la causa de muerte de una persona, se utiliza una prueba de independencia, empleando como metodología la realización de tablas de contingencia, la prueba χ^2 y las mediciones del p-valor y de la V de Cramer. Los resultados obtenidos en este análisis de independencia determinaron que múltiples variables son dependientes de las causas de la muerte. Se concluye que la investigación revela una relación significativa entre factores demográficos y causas de muerte en Costa Rica, utilizando pruebas estadísticas para identificar patrones de dependencia y proporcionando evidencias sólidas y útiles para políticas públicas.

METODOLOGÍA

Para toda esta investigación se utilizó el software R, en la versión 4.3.1, además, se utilizaron múltiples paquetes como *readxl* en su versión 1.4.3 (Wickham & Bryan, 2023), *dplyr* en la versión 1.1.4 (Wickham et al, 2023), *ggplot2* en la versión 3.5.1 (Wickham, 2016), *cowplot* en su versión 1.1.3 (Wilke, 2024), *rcompanion* en la versión 2.4.36 (Mangiafico, 2024) y *knitr* en la versión 1.47 (Xie, 2024).

Es fundamental definir lo que es una tabla de contingencia, ya que es el pilar para la prueba de χ^2 . “El análisis de tablas de contingencia (ATC) es una técnica destinada al estudio de la relación entre dos o más variables cualitativas o categóricas, es decir, medidas a nivel nominal y ordinal” (López-Roldán & Fachelli. 2015).

En el artículo de De la Fuente Fernández (2016), se describe el proceso para trabajar sobre una tabla de contingencia. Este autor empieza creando la hipótesis nula H_0 , y a su vez, define un nivel de significancia, el cual busca diferenciar cuándo se puede aceptar la hipótesis nula y cuándo se puede rechazar. Posteriormente, se describe en qué casos es posible usar la distribución χ^2 antes de entrar propiamente a la tabla de

contingencia. Además, se expone cómo hacer una tabla de contingencia y cuáles son los requisitos para usar la χ^2 . Por último, también se da una breve explicación de cómo interpretar los datos de una tabla de contingencia.

De esta manera, las tablas de contingencia son herramientas estadísticas utilizadas para analizar la relación entre dos variables categóricas. Son fundamentales en el análisis de datos, y se emplean para analizar la homogeneidad, dependencia e independencia entre variables categóricas. Más específicamente: “Cuando se desea comparar dos caracteres (X, Y) en una misma población que admiten las modalidades: $X(x_1, x_2, \dots, x_k)$, y $Y(y_1, y_2, \dots, y_p, \dots, y_m)$, se toma una muestra de tamaño n , representando por n_{ij} el número de elementos de la población que presentan la modalidad x_i de X e y_j de Y ” (De la Fuente Fernández, 2016, p. 4).

Complementando esta sección, otra herramienta útil es la V de Cramer. Esta es una medida para determinar cómo se asocian dos variables categóricas y evaluar así la fuerza de la relación entre ellas en tablas de contingencia. Se deriva del estadístico χ^2 , y se ajusta según el tamaño de la tabla. La fórmula para calcular la V de Cramer es:

$$V = \sqrt{\frac{\frac{\chi^2}{N}}{\min(k-1, r-1)}}$$

donde: χ^2 es el valor del estadístico chi-cuadrado, N es el tamaño de la muestra, k es el número de columnas y r es el número de filas.

La V de Cramer varía entre 0 y 1, donde 0 indica ninguna asociación y 1 indica una asociación perfecta (Betancourt & Caviedes. 2018).

Para finalizar, se define cómo se va a usar la prueba χ^2 para la presente investigación. Las pruebas de independencia a través de la χ^2 son un tipo de prueba de hipótesis, por lo que se debe definir una hipótesis nula. Para este caso es “Las distribuciones empíricas X e Y son independientes” (De la Fuente Fernández, 2016, p. 6). Así, se puede rechazar la hipótesis nula en caso de que:

$$\sum_{i=1}^k \sum_{j=1}^m \frac{(n_{ij} - e_{ij})^2}{e_{ij}} > \chi^2_{(1-\alpha), (k-1) \cdot (m-1)}$$

en donde: k representa el número de filas, m el número de columnas, n_{ij} es una entrada de la tabla, e_{ij} es la frecuencia esperada, $(k - 1) \cdot (m - 1)$ representa los grados de libertad de la χ^2 y α representa el nivel de significancia (De la Fuente Fernández, S., 2016, p.3-5).

RESULTADOS

Las pruebas de independencia entre el grupo de muerte creado por los estudiantes y las variables: región del Ministerio de Salud pública, nacionalidad, sexo, estado civil, ocupación y rango etario, resultaron en una clara dependencia. El análisis se comparó con el cuartil 90 de la χ^2 y el resultado fue mayor, lo que indica que la hipótesis nula debe ser rechazada.

De esta manera, la siguiente tabla resume los resultados obtenidos de las pruebas de independencia realizadas con las variables mencionadas.

Cuadro 1

Resultados de pruebas de independencia

	Grados de libertad	p-valor	V de Cramer
Región del MSP	32	$2,741949 \cdot 10^{-220}$	0,0392
Nacionalidad	4	$3,039234 \cdot 10^{-148}$	0,0609
Sexo	4	0	0,1087
Estado civil	16	0	0,1148
Ocupación	32	0	0,1472
Rango etario	64	0	0,2272

El cuadro anterior muestra un p-valor de 0 en las pruebas χ^2 realizadas con las variables sexo, estado civil, ocupación y rango etario, mientras que el p-valor obtenido de dichas pruebas con las variables región del MSP y nacionalidad fue prácticamente nulo, lo que significa que, si se acepta la hipótesis nula (independencia), la probabilidad de que sea cierta es nula o básicamente nula.

Sumado a lo anterior, el cuadro mostrado también presenta el valor de la V de Cramer para cada una de las variables analizadas. Este valor indica el nivel de asociación que existe entre el grupo de muerte creado por los estudiantes y cada una de las variables en la primera columna, con valores desde 0, el cual es el mínimo posible, hasta 1, el máximo posible.

En la tabla se agregan los grados de libertad del estadístico teórico χ^2 , esto debido a que influyen directamente en la interpretación del valor de la V de Cramer, pues, a mayor cantidad de grados de libertad, se requiere de un valor menor en la V de Cramer para afirmar una asociación media o grande (Kim, H. 2017. párrs 9-11).

Con esto, se destaca principalmente la variable de Rango etario, ya que es la que tiene el estadístico teórico con más grados de libertad y mayor valor de la V de Cramer, por lo que se puede afirmar que presenta una asociación bastante significativa con el grupo de muerte (Kim, H. 2017. párrs 9-11).

De forma similar con la ocupación, pues presenta un valor significativo a pesar de su gran cantidad de grados de libertad, lo cual indica que hay, al menos, una relación media entre esta variable y el grupo de causa de muerte.

Un caso distinto es con el sexo, que, según su valor de la V de Cramer, tiene una asociación baja con el grupo de causa de muerte, siempre basándose en la tabla de Kim, H. (2017, párrs 9-11).

También cabe mencionar que al realizar las pruebas de independencia con las variables de cantón de residencia y de muerte, resultaron en dependencia. Sin embargo, al ser variables tan numerosas, hacen que las tablas de contingencias tengan una gran cantidad de columnas, por lo que la prueba de independencia tiende a fallar, debido a que se ocupa una cantidad significativa de observaciones en cada casilla de la tabla para su correcto funcionamiento. Por lo tanto, el problema con las observaciones no permite dar una conclusión certera, por lo que no se puede afirmar si realmente existe una dependencia.

CONCLUSIONES

En el presente trabajo, se analiza la dependencia entre factores asociados a las causas de muerte en Costa Rica, buscando determinar, mediante pruebas estadísticas, si existe una relación significativa entre la causa de muerte y variables como la nacionalidad, el estado civil, la ocupación y el sexo de las personas fallecidas. La pregunta de investigación se adoptó para comprender mejor las tendencias en las tasas de mortalidad y las características demográficas influyentes, con el objetivo de proporcionar una base empírica para posibles intervenciones de salud pública. Los resultados confirmaron las expectativas iniciales, identificando patrones de dependencia considerables entre las variables analizadas y posibles incrementos en las tasas de mortalidad.

La metodología incluyó un análisis descriptivo de los datos de mortalidad del Instituto Nacional de Estadística y Censos (INEC) y el cálculo de tasas de mortalidad ajustadas por cada 1000 habitantes, utilizando estadísticas anuales de defunciones y poblaciones. Se emplearon pruebas estadísticas como la prueba χ^2 , junto con tablas de contingencia para explorar las relaciones entre los grupos de causa de muerte y distintas variables, proporcionando un marco claro para el análisis realizado. Los resultados obtenidos de la metodología empleada en la investigación se logran apreciar en el Cuadro 1, donde se indican los valores resultantes de las pruebas χ^2 y de la V de Cramer. A partir de este cuadro se puede concluir que las variables con las que se hizo el análisis de independencia son, ciertamente, dependientes respecto a la causa de muerte; además, con la información obtenida a partir del cuadro, se puede incentivar el descubrimiento del porqué estas variables influyen en la muerte y así en el futuro realizar una investigación más sólida y profunda.

Por lo tanto, el estudio realizado aborda un vacío significativo en la literatura al proporcionar un análisis detallado y actualizado de las tasas de mortalidad en Costa Rica, combinando datos demográficos con el impacto de la pandemia de COVID-19. Al hacerlo, ofrece nuevas perspectivas sobre cómo las características demográficas pueden impulsar el descubrimiento del porqué estas variables influyen en la mortalidad, mejorando así el entendimiento de los factores que afectan la salud pública en el país.

Durante el desarrollo del estudio se encontraron algunos problemas metodológicos, como la posible influencia sesgada de la pandemia de COVID-19 en los datos de los últimos años. Además, surgieron problemas como la inclusión de metodologías no utilizadas, como la prueba de Kolmogórov-Smirnov, que causó sobrecarga de información.

En este sentido, para futuras investigaciones, hay ciertas recomendaciones que podrían ser tomadas en cuenta. En primer lugar, profundizar en el análisis de otros factores demográficos, ampliando el estudio para incluir variables adicionales como el nivel educativo, los ingresos económicos o el acceso a servicios de salud. Además, utilizar métodos estadísticos más avanzados, en caso de necesitarse para un estudio más especializado.

Adicionalmente, un añadido a la investigación podría ser realizar estudios longitudinales, ya que estudiar cohortes específicas de población a lo largo del tiempo puede ayudar en gran medida a entender mejor las tendencias y causas de mortalidad. Sumado a lo anterior, se podría analizar el impacto de políticas de salud pública, evaluando cómo diferentes intervenciones y políticas de salud han influido en las tasas de mortalidad.

Para finalizar, mediante las metodologías estadísticas desarrolladas, el estudio revela la existencia de una relación entre factores demográficos y causas de muerte, proporcionando evidencias sólidas y útiles para políticas públicas y estrategias de salud. Con lo cual se evidencia que la causa de muerte es dependiente de la región de residencia del Ministerio de Salud pública, la nacionalidad, el sexo, el estado civil, la ocupación y el rango etario.

AGRADECIMIENTOS

Se agradece profundamente al tutor, PhD. Maikol Solís, por su orientación, apoyo y constante aliento durante todo el proceso de investigación. La calidad de sus consejos no solo mejoró el proyecto, sino que también impulsaron un mejor desarrollo en nuestras capacidades y conocimientos académicos. Estamos muy agradecidos por su dedicación, paciencia y tiempo brindado.

BIBLIOGRAFÍA

Betancourt, A., & Caviedes, I. (2018). Metodología de correlación estadística de un sistema integrado de gestión de la calidad en el sector salud. <https://dialnet.unirioja.es/descarga/articulo/6726339.pdf>

De la Fuente Fernández, S. (2016). Aplicaciones de la chi-cuadrado: tablas de contingencia, homogeneidad dependencia e independencia. Madrid: Universidad Autónoma de Madrid. <https://www.fuenterrebollo.com/Aeronautica2016/contingencia.pdf>

Instituto Nacional de Estadística y Censos, Área de Censos y Encuestas (2016), Unidad de Estadísticas Demográficas. (2023). Costa Rica: Total diseases 2014 - 2021. Kaggle.

<https://doi.org/10.34740/KAGGLE/DSV/6211711>

Kim, H. (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. The Korean Academy of Conservative Dentistry. <https://doi.org/10.5395/rde.2017.42.2.152>

López-Roldán & Fachelli, S. (2015). Análisis de tablas de contingencia. https://ddd.uab.cat/pub/caplli/2015/131469/metinvsoccuan_cap36a2015.pdf

Mangiafico, S. S. (2024). *rcompanion: Functions to Support Extension Education Program Evaluation* [version 2.4.36]. Rutgers Cooperative Extension. New Brunswick, New Jersey. <https://CRAN.R-project.org/package=rcompanion/>

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>

Wickham, H., & Bryan, J. (2023). *readxl: Read Excel Files* [R package version 1.4.3]. <https://CRAN.R-project.org/package=readxl>

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation* [R package version 1.1.4]. <https://CRAN.R-project.org/package=dplyr>

Wilke, C. O. (2024). *cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'* [R package version 1.1.3]. <https://CRAN.R-project.org/package=cowplot>

Xie, Y. (2024). *knitr: A General-Purpose Package for Dynamic Report Generation in R* [R package version 1.47]. <https://yihui.org/knitr/>

ANEXOS

Se adjunta el código de programación utilizado en caso de desear ver el proceso completo del trabajo:

https://github.com/Santifer0803/Proyecto_estadistica_1