

Técnicas estadísticas para identificar posibles relaciones bivariadas

Statistical techniques for possible identification on bivariate relations

Nelsa María Sagaró del Campo^{1*} <http://orcid.org/0000-0002-1964-8830>

Larisa Zamora Matamoros² <http://orcid.org/0000-0003-2210-0806>

¹Universidad de Ciencias Médicas, Santiago de Cuba. Cuba.

²Universidad de Oriente, Santiago de Cuba. Cuba.

*Autora para la correspondencia: nelsa@infomed.sld.cu

RESUMEN

Introducción: La mayoría de los problemas en la investigación biomédica son de naturaleza causal. El análisis en estos estudios debe comenzar por la búsqueda de asociación entre las variables que representan la causa y el efecto y solo si la asociación es significativa, continuará el análisis de inferencia causal.

Objetivo: Sistematizar las diferentes técnicas estadísticas que verifican una relación bivariada según el tipo de variable.

Métodos: Se realizó una revisión bibliográfica exhaustiva sobre el tema en las bases de datos biomédicas alojadas en la Internet. Se organizó el contenido por subtemas y se elaboró un material con una síntesis crítica de los aspectos más importantes, en el cual se plasmó además la experiencia de las autoras.

Resultados: Se expone, según tipo de variables, información básica de los coeficientes, pruebas de hipótesis y gráficos empleados en cada caso, las medidas de asociación para estudiar el riesgo, los atributos que aseguran la validez de una asociación, el azar y el sesgo como los errores que pudieran cometerse en el proceso de investigación y que pueden invalidar la existencia de una asociación. También se presenta la forma de analizar la asociación en el análisis estadístico implicativo.

Conclusiones: El conocimiento de los estadísticos para verificar una relación entre variables y la selección de técnicas estadísticas es esencial para llevar a cabo el proceso inicial de inferencia causal.

Palabras clave: técnicas estadísticas; relaciones bivariadas, asociación; correlación; odds ratio; análisis estadístico implicativo; causalidad en medicina.

ABSTRACT

Introduction: Most of the problems in biomedical research are of causal nature. The analysis of these studies should begin with the search for an association between the variables that represent the cause and the effect, and only if the association is significant will the causal inference analysis continue.

Objective: To systematize the different statistical techniques that verify a bivariate relationship according to the type of variable.

Methods: An exhaustive bibliographic review on the subject was carried out in the biomedical databases hosted in the Internet. The content was organized by sub-topics and a material with a critical synthesis of the most important aspects was elaborated, in which the experience of the authors was also expressed.

Results: According to the type of variables, we have presented basic information about the coefficients, hypothesis tests, and graphs used in each case, the association measures to study risk, the features that ensure the validity of an association; chance and bias are also exposed as the mistakes that could be made in the investigation process and that could invalidate the existence of an association. The way of analyzing the association in the implicative statistical analysis is also presented.

Conclusions: The knowledge of statisticians to verify a relationship between variables and the selection of statistical techniques is essential for carrying out the initial process of causal inference.

Keywords: statistic techniques; bivariate relations; association; correlation; odd ratios; implicative statistical analysis.

Recibido: 03/09/2019

Aprobado: 14/11/2019

Introducción

La mayoría de los problemas de investigación en las ciencias biomédicas son de naturaleza causal. El primer paso en el estudio de la causalidad es buscar la posible asociación entre las variables que se consideran causa y efecto, y solo si se encuentra asociación significativa entre ellas podrá continuar el análisis de inferencia causal.⁽¹⁾

Dos variables se relacionan entre sí cuando el cambio de una categoría a otra (para variables cualitativas) o de un valor a otro (para variables cuantitativas), provoca una modificación en las categorías o valores de la otra variable; en caso contrario se dirán independientes.

El análisis estadístico que identifica una posible relación entre variables es una forma básica de análisis de datos y según el tipo de variable puede llamarse asociación o correlación. Suele llamarse asociación cuando se busca la relación entre variables categóricas, de manera que sus categorías puedan colocarse en tablas de contingencia, y correlación cuando se relacionan variables en escala de intervalo o de razón o incluso, en escala ordinal y en vez de analizarla por tablas de contingencia, se les asignan rangos a las categorías convirtiéndose en numéricas.

Estas relaciones entre las variables pueden ser:

1. Relaciones simples: entre dos o tres variables aleatorias en la cual se establece:
 - a. Relación bivariada: relación entre dos variables (X, Y) donde X y Y son medidas en cualquier escala.
 - b. Relación bivariada con influencia de una tercera variable que puede ser:
 - Relación transicional: cuando la variable Y, dependiente en una relación bivariada (X, Y) es la independiente respecto a otra relación bivariada (Y, Z) $[X \rightarrow Y \rightarrow Z]$.
 - Interferencia de una o varias variables intervinientes: cuando una tercera variable (Z) provoca cambios importantes en la relación (X, Y).
 - Correlación parcial: correlación entre dos variables eliminando los efectos de otra variable interviniente.
2. Relaciones múltiples: cuando se correlacionan múltiples variables aleatorias, generalmente existe una variable aleatoria dependiente y un vector aleatorio independiente, formado por las covariables e incluye variantes como:
 - a. Correlación semiparcial: que permite conocer las contribuciones de las covariables sobre la variable dependiente eliminando el efecto de una o más covariables, según convenga.
 - b. Correlaciones canónicas: cuando se correlacionan dos vectores aleatorios.

En una relación entre variables se examina la intensidad por el valor del estadígrafo y el sentido por su signo. Además, para conocer la significación de la relación siempre se deben realizar contrastes de hipótesis. En el caso de las variables cuantitativas se examina, además, la forma en que se relacionan, la cual puede observarse gráficamente en un plano cartesiano o puede ser expresada matemáticamente. Esta expresión es importante para explicar e incluso, predecir lo que pasará en la variable dependiente en función de los cambios en la independiente.

Existe una gran variedad de coeficientes y pruebas paramétricas y sobre todo, no paramétricas para corroborar la significación de una relación encontrada, estos se emplean en dependencia del tipo de variables y la cantidad y tipo de muestra si fuera el caso.⁽²⁾

Por la importancia que tiene para todo investigador el conocimiento de estas técnicas se gestó este trabajo con el objetivo de sistematizar las diferentes técnicas estadísticas para verificar la relación entre variables según su tipo.

Métodos

En este estudio se abordan solo las relaciones bivariadas clásica y con la influencia de una tercera variable. La investigación se desarrolló con una exhaustiva revisión de la literatura sobre el tema en las bases de datos bibliográficas alojadas en la Internet como Pubmed/Medline, EMBASE, SCOPUS, Web of Science, EBSCOhost y Google Académico, en el período entre enero y abril de 2019. Se emplearon en la estrategia de búsqueda como descriptores los consignados en las palabras claves del artículo, estos fueron utilizados en idiomas inglés, francés, portugués y español, combinados con los operadores lógicos. Se escogieron los artículos publicados en revistas arbitradas o repositorios de Universidades, cuyo período de publicación fuera, en lo posible, menor de 10 años y se relacionaran con la medición en salud. Se organizó el contenido por subtemas y se elaboró un material con una síntesis crítica de los aspectos más importantes, en el cual se plasmó, además, la experiencia de las autoras.

Técnicas estadísticas según tipo de variables

Se expone a continuación, información básica de los coeficientes, pruebas de hipótesis y gráficos empleados, según tipo de variables. Las fórmulas y otros detalles de los coeficientes expuestos se pueden consultar en las bibliografías citadas.

Relación entre variables cuantitativas

La relación entre dos variables cuantitativas puede medirse por la covariación, pero esta depende de las unidades de medida de las variables y no está acotada, por lo que se prefiere el uso de los coeficientes de correlación que permiten medir la fuerza y la dirección de la asociación entre ambas variables. Entre estos están:

- El coeficiente de correlación producto-momento de *Pearson*: coeficiente paramétrico que solo puede calcularse para variables con niveles de medición intervalar o de razón.⁽³⁾
- Los coeficientes de correlación no paramétricos Rho de *Spearman* y Tau de Kendal o por rangos de *Kendall*: estos no requieren para su empleo que la distribución muestral ajuste a una distribución conocida, por lo que los estimadores muestrales no son representativos de los parámetros poblacionales. Se emplean también para variables medidas en escala ordinal. Se calculan en base a una serie de rangos asignados por lo que no se afectan por valores atípicos y son apropiados para relaciones asimétricas, no lineales.

Los valores de estos coeficientes oscilan de -1 a 1 , siendo los valores extremos los que indican la mayor correlación y el 0 la ausencia de correlación. El signo del coeficiente indica el sentido de la relación. Ante un signo positivo se dirá que la relación es directa (las variables cambian en el mismo sentido) y ante uno negativo se dirá que es inversa (a medida que aumenta una disminuye la otra).^(4,5)

La interpretación del valor del coeficiente puede darse con mayor detalle por intervalos. Diferentes son los intervalos que han establecido algunos autores para considerar la correlación débil, moderada o fuerte, lo cual aplica igual en el caso de un valor negativo. Incluso ciertas literaturas incluyen intervalos no excluyentes como esta guía de la Universidad de Chile de Ciencias Sociales que plantea que un valor entre 0 y $0,2$ indica una correlación mínima; entre $0,2$ y $0,4$, una correlación baja; entre $0,4$ y $0,6$, moderada, entre $0,6$ y $0,8$ buena y entre $0,8$ y 1 muy buena.⁽⁵⁾

Es importante tener presente algunos aspectos en el estudio de la correlación entre variables:

- La correlación no implica causalidad. La causalidad es un juicio de valor que requiere más información.
- No necesariamente existirá dependencia entre ambas variables, ya que una correlación puede deberse a una covariación conjunta determinada por una tercera variable
- Estos coeficientes solo miden correlación monótona y lineal. Para medir una relación curvilínea se emplea la proporción de correlación representada por el coeficiente eta.⁽⁶⁾
- Tampoco son apropiados para medir concordancia, en este caso se emplea el método gráfico de *Bland y Altman* o el coeficiente de correlación intraclase que estima el promedio de las correlaciones entre todas las posibles ordenaciones de los pares de observaciones disponibles.⁽⁷⁾ Ambos métodos pueden interpretarse como una medida de fiabilidad.^(8,9)

El gráfico más adecuado para apreciar la relación entre dos variables numéricas es el diagrama de dispersión. Este que consiste en plotear los pares de valores de las variables de cada unidad de observación en el plano cartesiano, formando una nube de puntos que puede adoptar diferentes formas las cuales ofrecen idea del tipo de relación (asemeja una recta cuando existe correlación lineal o una curva para una relación curvilínea) o sin forma específica que indica la independencia entre las variables.⁽³⁾

En el caso de darse una correlación lineal con dependencia entre las variables es posible estimar la recta que mejor describe esta relación entre ambas variables mediante la regresión lineal simple y mediante el coeficiente de determinación que evaluará el ajuste de los datos a la recta obtenida.⁽¹⁰⁾

Relación entre variables ordinales

Una relación entre dos variables ordinales se da cuando los cambios en el orden de las categorías de una variable influyen en el orden de las categorías de la otra variable. En este

caso se emplean coeficientes no paramétricos que trabajan con rangos y se basan en el concepto de inversión y no-inversión.⁽¹¹⁾ Una relación positiva o de “no inversión” habla de un predominio de asociación entre los rangos altos de una variable con los rangos altos de la otra al igual que entre los rangos bajos de ambas variables. En cambio, una relación negativa (inversión) habla de un predominio de asociación entre los rangos altos de una variable con los bajos de la otra. Si las variables son independientes habrá tantas inversiones como no inversiones y el coeficiente valdrá cero; sin embargo, un valor de cero no implica necesariamente que sean independientes.⁽⁴⁾

Entre estos coeficientes están:

- Rho de Spearman empleado para comparar dos conjuntos de rangos ordenados en una muestra o dos grupos con los rangos ordenados de varias unidades de análisis.⁽¹²⁾
- Tau b de Kendall empleado para comparar dos rangos cuando se tiene un par de rangos por cada unidad de observación. Es solo aplicable en tablas de contingencia cuadradas (igual número de filas que de columnas) y si ninguna frecuencia marginal tiene valor cero en sus casillas; no llega a valer 1 si la tabla no es cuadrada.⁽¹²⁾
- Tau c de Kendall que supera las dificultades del tau b.
- Gamma de Goodman y Kruskal.
- d de Somers. Este tiene una versión simétrica que coincide con el Tau-b de Kendall y una asimétrica, modificación del gamma, que considera a las variables como dependiente e independiente.⁽¹²⁾

El Gamma se emplea para identificar la relación entre dos variables cuando al menos una variable es ordinal. Spearman y Kendall pueden emplearse cuando ambas son ordinales o en escala intervalar.

Todos estos coeficientes arrojan resultados entre -1 y 1, excepto la versión asimétrica del d de Sommer.⁽¹²⁾ Todos se basan en los conceptos de inversión y no inversión, la diferencia entre ellos está en el tratamiento que dan a los empates (cuando los pares no son de tipo inversión ni no inversión ya que los rangos en ambas variables coinciden).

Para conocer la concordancia entre dos o más observadores es posible emplear una versión del coeficiente de Kendall y el W de concordancia de Kendall.

Relación entre una variable continua y una nominal

Las técnicas empleadas para estudiar la relación entre una variable continua y una nominal varían en dependencia de la variable cualitativa nominal y las características de las muestras en estudio.

Si la variable cualitativa es politómica se pueden emplear el coeficiente de correlación esta que no supone linealidad y cuyo cuadrado puede interpretarse, si el diseño lo permite, como la proporción de varianza de la variable cuantitativa que es explicada por la variable categórica⁽¹⁵⁾ o el índice f de Cohen cuyo valor mide la intensidad de la asociación.⁽³⁾

En cambio, si la variable es dicotómica se podrán emplear el coeficiente biserial puntual, los índices delta de Glass, g de Hedges, d de Cohen o el índice d, modificación del índice d de Cohen para el caso de una medición repetida en dos momentos para un mismo grupo.⁽³⁾

Estos índices son muy empleados para determinar el llamado tamaño del efecto que se produce en grupos bajo diferentes tratamientos en estudios experimentales o en el metaanálisis de investigaciones cuyo efecto cuantitativo se obtuvo por la media.⁽¹³⁾

A veces las categorías de una variable cualitativa se forman a partir del empleo en la investigación de más de una muestra, en este caso lo que se pretende es buscar diferencias entre ellas. Si las diferentes categorías de la variable politómica representan los niveles donde se mide la variable cuantitativa se debe emplear el análisis de varianza paramétrico o el no paramétrico de Kruskal Wallis. Dada una dicotomía condicionada por dos muestras independientes se pueden emplear las pruebas de comparación de media en muestras independientes t de student sin o con la aproximación de Welch, en dependencia de si existe o no homocedasticidad de varianzas, respectivamente; o sus alternativas no paramétricas, la U de Mann y Withney, las rachas de Wald Wolfowitz, entre otros. Si la dicotomía es condicionada por muestras pareadas se debe emplear la t de student para muestras pareadas o sus alternativas no paramétricas el test de los signos, el de rangos con signos de Wilcoxon, entre otros.⁽⁶⁾

El empleo de una prueba paramétrica siempre dependerá del cumplimiento de supuestos como la independencia, normalidad y homocedasticidad de las varianzas.⁽¹⁴⁾ El supuesto de independencia puede corroborarse mediante el test de Durbin Watson, el de normalidad a través de técnicas gráficas como el histograma, los gráficos de probabilidad normal P-P y de cuantiles normales Q-Q o con técnicas estadísticas como las pruebas de bondad del ajuste de Kolmogorov-Smirnov, Shapiro-Wilk, D'Agostino, Lilliefors, entre otros o el cálculo de los coeficientes de simetría y de curtosis, mientras que la igualdad de varianzas se verifica con los test de Levene, Bartlett o C de Cochran.

Los gráficos empleados para representar esta información son los polígonos de frecuencia, donde se podrá comparar el grado de solapamiento de las distribuciones de la variable cuantitativa condicionadas por las categorías de la cualitativa y en el análisis exploratorio de los datos se emplean las gráficas de cajas y bigotes o las de barras de error múltiples con la distribución de la variable cuantitativa condicionada a la variable cualitativa (una caja o barra por cada categoría de la variable nominal).⁽¹⁵⁾

Relación entre variables cualitativas

Para identificar la posible relación entre dos variables cualitativas hay que observar si la distribución de las categorías de una de las variables difiere en función de las de la otra, o sea comparar las distribuciones condicionadas de una de las dos variables agrupadas en función de los valores de la otra. Si no hay relación entre las variables estas distribuciones deberían ser iguales.⁽¹⁵⁾

Habitualmente se colocan los datos en una tabla de contingencia o de doble entrada, donde aparecen las frecuencias observadas (frecuencias absolutas conjuntas o número de casos que presentan simultáneamente las modalidades fila y columna) y se emplean métodos directos como el análisis de los residuos de la diferencia entre valores observados y esperados o la descomposición de la tabla en tablas de 2×2 .⁽¹²⁾

Esta relación se mide a través de los llamados coeficientes de contingencia tales como:

- Coeficientes basados en la reducción de error como Lambda, Tau de Goodman y Kruskal y el coeficiente de incertidumbre de Theil.⁽²⁾

- Coeficientes simétricos basados en chi cuadrado como:
 - El T de Tschuprow.
 - El de contingencia C de Pearson (solo para tablas cuadradas inicialmente y corregido por Pawlik superando esta dificultad).⁽²⁾
 - Si al menos una de ellas es binaria se puede emplear el coeficiente de contingencia V de Cramer.
 - En el caso que ambas sean dicotómicas se emplea el coeficiente Phi de Pearson.

Para las tablas de 2×2 existen muchísimos más coeficientes como el Q de Yule, J de Youden, de Forbes, de McConnaughy, de Michel, de Simpson, de Baroni-Urbani y Buser, de Agrell e Inverson, de Mountford, de Dice, entre otras.⁽¹⁶⁾

Si se quiere determinar la concordancia entre instrumentos de medida cuyo resultado es categórico se emplea el índice de Kappa de Cohen que depende del acuerdo observado, pero también de la prevalencia del carácter estudiado y de la simetría de los totales marginales.⁽¹⁷⁾

Para demostrar la significación de esta asociación solo se emplean pruebas no paramétricas como las siguientes:

- Chi cuadrado.
- G^2 de razón de verosimilitud.
- Para tablas de 2×2 puede emplearse el test exacto de probabilidades de Fisher, sobre todo en muestras menores de 20 y válido para variables ordinales también si la muestra es pequeña.⁽¹⁸⁾ Este aumenta su potencia con la modificación propuesta por Tocher.
- Si la dicotomía es producto de muestras pareadas se emplea la prueba de McNemar.⁽¹⁸⁾
- Si se quiere controlar el efecto de una tercera variable, se analizan ambas variables dicotómicas dentro de los estratos de esa tercera, mediante la prueba de Cochran Mantel Haenszel.

Una mención aparte merece la más empleada de estas técnicas, la chi cuadrado, llamada de independencia cuando la tabla se genera a partir de una muestra o de homogeneidad si se genera por más de una muestra y se fijan entonces esos totales marginales por fila (estudios de casos y controles) o por columna (estudios de cohortes o ensayos clínicos controlados). Esta prueba solo mide la significación de la asociación, pero no constituye una medida de intensidad de la asociación, porque en su cálculo involucra el tamaño de muestra (depende no solo de las proporciones totales sino también del gran total), tampoco permite conocer el sentido de la asociación pues al estar el estadígrafo elevado al cuadrado su valor siempre será positivo. El test no se puede emplear si en más de 20 % de las celdas de la tabla hay valores esperados menores que 5 o en una sola celda hay un valor menor que 1. Para aumentar estas frecuencias se sugiere agrupar las categorías siguiendo un sentido lógico siempre que sea posible o emplear el test de Fisher en tablas de 2×2 . Otro aspecto a considerar en las tablas de 2×2 es el empleo de la corrección de Yates atenuando el efecto de usar una distribución continua para representar una distribución discreta y evitar encontrar diferencias muestrales que no existen a nivel poblacional (dificultar el rechazo de la hipótesis nula).⁽¹¹⁾

La representación gráfica de la relación entre variables categóricas puede obtenerse mediante barras múltiples o barras proporcionales. En las barras se deben representar las frecuencias relativas condicionadas y no las frecuencias absolutas, pues así se elimina el efecto de la diferencia de tamaño entre los subgrupos.⁽³⁾

Validez de una asociación

Para que una asociación sea válida (real o verdadera) hay que tener en cuenta el efecto de dos errores que pudieran cometerse en el proceso de investigación: el azar (error aleatorio) y el sesgo (error sistemático).

La influencia del azar se puede determinar por dos vías:

- Las pruebas de significación estadística, donde la significación está dada por el valor de p . Si una prueba de asociación arroja como resultado un valor de p mayor que el nivel de significación alfa prefijado, habitualmente de 5 %, no es posible asegurar que existe

asociación entre las variables, aun cuando el valor del coeficiente así lo indique, porque la probabilidad de estar equivocado al señalar que existe asociación es muy alta.⁽¹⁹⁾

- Los intervalos de confianza, como por ejemplo, al 95 % reflejan los resultados que se esperarían obtener en 95 de 100 repeticiones del mismo estudio, por lo que no excluyen la posibilidad de que exista dicha asociación y es posible que proporcionen cierta información sobre su magnitud, en especial en estudios pequeños.⁽²⁰⁾

Ambas informaciones son complementarias y dependientes del tamaño de la muestra, de manera que, a menor tamaño de muestra, mayor es el valor p y más amplio el intervalo. Por otro lado, los sesgos pueden incidir en la obtención de una asociación que no existe o hacerla mayor de lo que es. Estos sesgos pueden ocurrir en cualquier fase del proceso de investigación y pueden ser de tres tipos.⁽²¹⁾

1. De selección: por una selección de los sujetos en estudio de manera diferencial entre grupos o de manera no representativa de la población que les da origen, por ejemplo: a) el sesgo de Berkson, que se produce en los estudio de casos y controles realizados en los hospitales, donde la probabilidad de ser incluido como caso es mayor, por lo que se producirá una sobrestimación de la asociación, b) pérdidas durante el seguimiento relacionadas con el factor de estudio o la variable de respuesta (pérdidas informativas), c) ausencia de respuesta que prevalece en relación con determinada variable, d) supervivencia selectiva que se produce cuando se seleccionan casos prevalentes en lugar de incidentes, típico en los estudios de casos y controles y e) el sesgo de autoselección por participación de voluntarios.
2. De medición: sucede durante la fase de obtención de información por empleo de criterios diferentes para la recolección de datos y clasificación de los grupos en estudio o cuando los métodos de medición para establecer las conclusiones se emplean de manera diferencial entre los grupos, por ejemplo en el uso de pruebas de detección de un efecto, cuya sensibilidad y especificidad varían en función de la exposición y el grado de asociación (que dependerá del número de resultados falsos positivos y negativos determinado por cada prueba). También se puede presentar el sesgo de memoria en el sujeto que no recuerda el dato o del entrevistador que influye en la respuesta.

3. De confusión: se produce cuando no se tiene en cuenta alguna variable responsable de la falsa asociación encontrada, que es el verdadero factor de riesgo pronóstico y es causa común del falso factor de riesgo o pronóstico y el desenlace.

El efecto del azar es imposible de controlar porque es inherente al muestreo, en cambio los sesgos se atenúan con un adecuado rigor metodológico en el estudio.

Medidas de asociación ligadas al riesgo de desarrollar un efecto frente a una exposición

La fuerza de asociación es la magnitud con que aumenta el riesgo de desarrollar un efecto cuando se presenta una exposición. A mayor fuerza de asociación es mayor la seguridad sobre la causalidad. Las tres medidas más empleadas en los estudios de causalidad en relación con este riesgo son:^(22,23)

1. Riesgo relativo (RR) que es la razón entre el riesgo de los expuestos (incidencia del efecto en los expuestos) y el riesgo de los no expuestos (incidencia del efecto en los no expuestos). Es la medida que refleja mejor la probabilidad de padecer la enfermedad en función de la exposición.

Esta medida es aplicable a los estudios prospectivos, ya sean experimentales (ensayos aleatorios) u observacionales (de cohorte), donde es posible estimar la incidencia del efecto. En los estudios de casos y controles se utiliza una estimación indirecta del riesgo relativo, a través de la razón de productos cruzados o razón de momios, más conocida como odds ratio y en los transversales la medida de asociación equivalente es la razón de prevalencias para enfermedades agudas y la razón de momios de prevalencia para enfermedades crónicas.

Un RR mayor de 1 tiene significación estadística y uno mayor de 3 tiene además significación clínica, pues implica un riesgo de desarrollar el efecto tres o más veces, mayor en los sujetos expuestos, que en los no expuestos.⁽²⁴⁾ *Silva*⁽²⁵⁾ plantea que para algunos autores un RR mayor que 2 es suficiente para considerar una relación fuerte.

2. Riesgo atribuible o diferencia de riesgos, la cual permite distinguir el efecto absoluto de la exposición y expresa la proporción de individuos expuestos que, por efecto de la exposición, desarrollarán el efecto. Consiste en restar al riesgo en expuestos, el riesgo en no expuestos. El resultado será la proporción de enfermedad que podrá disminuirse al eliminar el factor de riesgo. Se le reconocen dos limitaciones: 1) no puede estimarse en los estudios de caso y controles ya que solo puede utilizarse cuando se estima alguna medida de frecuencia como la prevalencia en un estudio transversal analítico o la incidencia en un estudio de cohorte y 2) puede subestimar la proporción de riesgo.
3. Fracción etiológica o porcentaje de riesgo atribuible, que permite estimar la proporción del riesgo o el efecto observado en los sujetos expuestos, que es atribuible a la exposición. Se puede estimar para los individuos expuestos denominado riesgo atribuible en expuestos y para la población denominado riesgo atribuible poblacional, a partir de la medida de asociación calculada según el tipo de estudio.

El riesgo atribuible y la fracción etiológica se consideran medidas del impacto potencial porque sirven para evaluar la efectividad de las acciones del personal de salud para atenuar o erradicar los supuestos factores de riesgo identificados tanto en los expuestos como en la población en general.⁽¹⁵⁾

Hace algunos años, era común emplear las tablas de 2×2 para organizar los datos procedentes de los estudios de causalidad, en la cual solo es posible analizar la relación entre dos variables. En sus columnas cuentan los individuos que poseen y no poseen la característica en estudio, generalmente la enfermedad o desenlace fatal y en sus filas, los expuestos y los no expuestos al supuesto factor de riesgo o pronóstico. Las medidas anteriores se pueden calcular a partir de los valores de las celdas de esta tabla. Sin embargo, en la actualidad, en correspondencia con el paradigma multicausal imperante se emplean técnicas de análisis multivariado y la medida de asociación se obtiene con la aplicación de la regresión logística binaria a partir del valor de la exponencial del coeficiente de regresión asociado a cada factor. Este análisis no solo permite fortalecer el argumento de la causalidad entre un factor de riesgo o pronóstico y un desenlace específico, además ayuda a un mejor reconocimiento y tratamiento de variables confusoras y modificadoras del efecto. En *Sagaró y Zamora*⁽²⁶⁾ se precisan detalles de la evolución histórica de las técnicas estadísticas para el

estudio de la causalidad donde se distingue el paso de las técnicas bivariadas a las multivariadas.

La asociación en el análisis estadístico implicativo

Las medidas que reflejan la relación entre variables en el análisis estadístico implicativo son:^(27,28,29,30,31,32)

- El índice de implicación: indicador de la no implicación de una variable a sobre otra variable b. Este índice es no simétrico y no coincide con el coeficiente de correlación u otros índices simétricos que miden asociación.
- El índice de implicación-inclusión o de implicación entrópica: versión entrópica del índice de implicación que supera la poca discriminación de este en muestras grandes. Este índice permite determinar el criterio entrópico al integrar la información a partir de la presencia de un escaso número de contraejemplos, tanto por la regla $a \Rightarrow b$ como por su negación $\neg b \Rightarrow \neg a$.

La representación gráfica de la relación se visualiza en el grafo implicativo, donde el sentido de la flecha orienta al investigador para generar hipótesis de causalidad. Este grafo es similar a los diagramas causales del tipo grafo acíclico dirigido que permiten plantear posibles rutas de asociación entre causas y efectos, así como otras rutas alternativas que pueden ocasionar sesgos por lo que facilitan también el control de posibles sesgos.^(33,34,35,36,37,38)

Conclusiones

El conocimiento de los estadísticos a emplear para verificar una relación entre variables es esencial para llevar a cabo el proceso inicial de inferencia causal. La selección de una prueba o coeficiente dependerá del tipo de variable, además el empleo de un test paramétrico conlleva el cumplimiento de supuestos como la normalidad, independencia y homocedasticidad de varianzas. Las pruebas no paramétricas se emplean para muestras

pequeñas, variables cualitativas o no cumplimiento de los supuestos. Independientemente de que se obtenga un valor alto en un coeficiente debe verificarse la significación del mismo, de lo contrario no es posible hablar de asociación y aun cuando se obtenga significación estadística esto no es siempre equivalente a una asociación en la práctica médica, esta conclusión solo puede lograrse con la lógica y experticia del investigador.

Otro aspecto en las relaciones bivariadas son las medidas específicas a emplear en estudios epidemiológicos y el control de errores potenciales desde el momento del diseño de la investigación.

Por último, se destaca que mediante el análisis estadístico implicative es posible obtener las relaciones entre variables de manera visual, lo cual pudiera resultar muy atractivo para el investigador.

Referencias bibliográficas

1. Castellanos Olivares A, Rojas Peñaloza J, Vásquez Márquez PI. ¿Cómo evaluar artículos de riesgo o causalidad en medicina? Taller: Metodología para aplicar la medicina basada en evidencias. Rev Mex Anest. 2016[acceso: 12/01/2019];39. Supl. 1: S232-5. Disponible en: <http://www.medigraphic.com/rma>
2. Flores Ruiz E, Miranda Novales MG, Villasís Keever MÁ. El protocolo de investigación VI: cómo elegir la prueba estadística adecuada. Estadística inferencial. Rev Alerg Mex. 2017[acceso: 10/01/2019];64(3):364-70. Disponible en: http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S2448-91902017000300364
3. Molina G, Rodrigo MF. Estadísticos de asociación entre variables. Estadística descriptiva en Psicología. Universitat Valencia Open CourseWare Curso 2009-2010. [acceso: 10/01/2019]. Disponible en: http://www.ocw.uv.es/ciencias-de-la-salud/pruebas-1/1-3/t_08-1.pdf
4. Camacho Sandoval J. Asociación entre variables cuantitativas: análisis de correlación. Acta Med Costarric. 2008[acceso: 03/01/2019];50(2). Disponible en: http://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S0001-60022008000200005

5. Díaz I, García C, León M, Ruiz F, Torres F, Lizama P, *et al.* Asociación entre variables (Pearson y Spearman en SPSS) Guía de Ayudantía Estadística Departamento de Sociología. Universidad de Chile Facultad de Ciencias Sociales (FACSO). 2014[acceso: 10/01/2019]. Disponible en: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwiBw5bq6KH0AhVLmeAKHdeOAW0QFjAAegQIBhAB&url=https%3A%2F%2Fwww.u-cursos.cl%2Ffacso%2F2014%2F2%2FSO01007%2F1%2Fmaterial_docente%2Fbajar%3Fid_material%3D994690&usg=AOvVaw3J6Vz9Lx1-iPVCLXck_vmX
6. Frey L, Botan C, Kreps G. Analyzing Relationships Between Variables. En: Investigating communication: An introduction to research methods. 2 ed. Boston: Allyn & Bacon; 1999.
7. Prieto L, Lamarca R, Casadop A. La evaluación de la fiabilidad en las observaciones clínicas: el coeficiente de correlación intraclase. Med Clin (Barc). 1998[acceso: 10/01/2019];110:142. Disponible en: <https://www.mvclinic.es/wp-content/uploads/Prieto-Coeficiente-correlaci%C3%B3n-intraclase.pdf>
8. SERGAS. Manual de Epidat 4. Ayuda de Concordancia y Consistencia. 2014[acceso: 10/01/2019]. Disponible en: http://www.sergas.es/gal/documentacionTecnica/docs/Saude Publica/Apli/Epidat4/Ayuda/Ayuda_Epidat4_Concordancia_y_consistencia_Octubre2014.pdf
9. Nayak BK, Haz A. How to choose the right statistical test? Indian J Ophthalmol. 2011[acceso: 10/01/2019];59(2):85-6. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3116565/>
10. Montanero Fernández J, Minuesa Abril C. Estadística básica para Ciencias de la Salud. Universidad de Extremadura. Servicio de Publicaciones Cáceres; 2018[acceso: 10/01/2019] Disponible en: <https://core.ac.uk/download/pdf/152461717.pdf>
11. De la Fuente Fernández S. Tablas de Contingencia. Análisis de variables categóricas. Facultad de Ciencias Económicas y Empresariales. UAM; 2011[acceso: 13/01/2019]; Disponible en: <http://www.estadistica.net/ECONOMETRIA/.../CONTINGENCIA/tablas-contingencia.pdf>

12. Akoglu H. User's guide to correlation coefficients. Turkish Journal of Emergency Medicine. 2018[citado: 10/01/2019];18(3):91-3. Disponible en: <https://doi.org/10.1016/j.tjem.2018.08.001>
13. Ledesma R, Macbeth G, Cortada de Kohan N. Tamaño del efecto: revisión teórica y aplicaciones con el sistema estadístico ViSta. Revista Latinoamericana de Psicología 2008[acceso: 10/01/2019];40(3):425-39. Disponible en: <http://publicaciones.konradlorenz.edu.co/index.php/rlpsi/article/download/358/257>
14. Jorge Dagnino S. Elección de una prueba de hipótesis. Rev Chil Anest. 2014[acceso: 10/01/2019];43:139-42. Disponible en: http://www.sachile.cl/upfiles/revistas/54e63841d7f47_13_eleccion-2-2014_edit.pdf
15. Jorge Dagnino S. Elección de una prueba de hipótesis. Rev Chil Anest. 2014[acceso: 10/01/2019];43:139-42. Disponible en: http://www.sachile.cl/upfiles/revistas/54e63841d7f47_13_eleccion-2-2014_edit.pdf
16. Rodríguez Salazar MN, Álvarez Hernández S, Bravo Nuñez E. Coeficientes de asociación. México: Ed. Plaza y Valdes SA; 2001.
17. Tang W, Hu J, Zhang H, Wu P, He H. Kappa coefficient: a popular measure of rater agreement. Shanghai Arch Psychiatry. 2015[acceso: 05/01/2019];27(1):62-7. DOI: 10.11919/j.issn.1002-0829.215010. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4372765/>
18. Gómez Gómez M, Danglot Banck C, Vega Franco L. ¿Cómo seleccionar una prueba estadística? (Segunda parte). Rev Mex Pediatr. 2013[acceso: 05/01/2019];80(2):81-5. Disponible en: <https://www.medigraphic.com/pdfs/pediat/sp-2013/sp131g.pdf>
19. Jiménez Paneque J. El valor de p en entredicho: significación estadística, clínica y práctica. Medwave. 2016[acceso: 10/01/2019];16(8):e6534 DOI: 10.5867/medwave.2016.08.6534 Disponible en: <http://www.medwave.cl/link.cgi/Medwave/Perspectivas/Cartas/5894>
20. Silva Ayçaguer LC. Confidence intervals and p values. Medwave. 2014[acceso: 05/01/2019];14(1):e5894. DOI: 10.5867/medwave.2014.01.5894 Disponible en: <http://www.medwave.cl/link.cgi/Medwave/Perspectivas/Cartas/5894>

21. Dawson Saunders B, Trapp RG. Obtención de inferencias a partir de datos: Intervalos de confianza y pruebas de hipótesis. En: Bioestadística Médica. 2 ed. México. El Manual Moderno; 1997. p. 99-118.
22. Szklo M. Epidemiologia translacional: algumas considerações. Epidemiol. Serv. Saúde. 2015[acceso: 15/01/2019];24(1):161-72. Disponible en: https://www.scielo.org/article/ssm/content/raw/?resource_ssm_path=/media/assets/ress/v24n1/2237-9622-ress-24-01-00161.pdf
23. Fajardo Gutiérrez A. Medición en epidemiología: prevalencia, incidencia, riesgo, medidas de impacto. Rev Alerg Mex. 2017[acceso: 10/01/2019];64(1):109-20. Disponible en: http://www.scielo.org.mx/scielo.php?pid=S2448-91902017000100109&script=sci_arttext
24. De Irala J, Martínez González MA, Guillén Grima F. ¿Qué es una variable de confusión? Med Clin (Barc). 2001;117(10):377-85.
25. Silva Ayçaguer LC. El acertijo de la causalidad. En: Cultura estadística e investigación científica en el campo de la salud: una mirada crítica. Madrid: Díaz de Santos; 1999.
26. Sagaró Del Campo NM, Zamora Matamoros L. Evolución histórica de las técnicas estadísticas y las metodologías para el estudio de la causalidad en ciencias médicas. MEDISAN. 2019[acceso: 12/04/2019];23(3) Disponible en: <http://www.medisan.sld.cu/index.php/san/article/view/2434>
27. Gras R. Genese et developpement de l'analyse statistique implicative: retrospective historique Educ Matem Pesq São Paulo. 2014[acceso: 11/01/2019];16(3):645-61. Disponible en: https://revistas.pucsp.br/index.php/emp/article/viewFile/21541/pdf_1
28. Gras R, Régnier JC, Lahanier Reuter D, Marinica C, Guillet F. L'Analyse Statistique Implicative. Des Sciences dures aux Sciences Humaines et Sociales 3ème édition revue et augmentée. Cépaduès Editions; 2017. ISBN: 978.2.36493.577.8
29. Gras. R, Couturier R, Gregori P. Un mariage arrangé entre l'implication et la confiance? VII International Conference SIA; June 2015[acceso: 14/01/2019]. Disponible en: https://www.researchgate.net/publication/284534810_Un_mariage_arrange_entre_l%27implication_et_la_confiance

30. Couturier R, Gregori P. Classification en utilisant les règles d'implication de l'ASI. Conference Paper; October 2017[acceso: 14/01/2019] Disponible en: <https://www.researchgate.net/publication/324770682>
31. Zamora Matamoros L, Díaz Silvera JR, Portuondo Mallet L. Fundamental Conceptson Classification and Statistic Implicative Analysis for Modal Variables Rev Colomb Estad. 2015[acceso: 10/01/2019];38(2):335-51. Disponible en: <http://dx.doi.org/10.15446/rce.v38n2.51664>
32. Zamora Matamoros L, Díaz Silvera J. Estudio de relaciones causales entre indicadores de promoción del primer año de la carrera Ciencia de la Computación con el análisis implicative modal. Revista Maestro y Sociedad. 2018[acceso: 14/01/2019];15(2):203-12. Disponible en: <https://revistas.uo.edu.cu/index.php/MyS/article/view/3520/3112>
33. Rodríguez Cortes T, Faerstein E, Struchiner CJ. Utilização de diagramas causais em epidemiologia: um exemplo de aplicação em situação de confusão. Cad Saúde Pública. 2016[acceso: 10/01/2019];32(8):e00103115. Disponible en: <http://dx.doi.org/10.1590/0102-311X00103115>
34. Brewer L, Wright J, Rice G, Neas L, Teuschler L. Causal inference in cumulative risk assessment: The roles of directed acyclic graphs. Environ Int. 2017;102: 30-41.
35. Greenland S, Pearl J. Causal diagrams. Wiley Stats Ref: Statistics Reference Online, John Wiley & Sons, Ltd. 2017. DOI: 10.1002/9781118445112.stat03732.pub2.
36. Werlinger F, Cáceres D. Aplicación de grafos acíclicos dirigidos en la evaluación de un set mínimo de ajuste de confusores: un complemento al modelamiento estadístico en estudios epidemiológicos observacionales. Rev Med Chile. 2018[acceso: 10/01/2019]; 146(7):907-13. Disponible en: <http://dx.doi.org/10.4067/s0034-98872018000700907>
37. Sagaró Del Campo NM, Zamora Matamoros L. Métodos gráficos en la investigación biomédica de causalidad. Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta. 2019[acceso: 20/01/2019];44(4). Disponible en: <http://revzoilomarinellosld.sld.cu/index.php/zmv/article/view/1846>

38. Sagaró Del Campo NM, Zamora Matamoros L. ¿Por qué emplear el análisis estadístico implicativo en los estudios de causalidad en salud? Revista Cubana de Informática Médica. 2019[acceso: 10/06/2019];11(1):88-103. Disponible en: http://revinformatica.sld.cu/index.php/rcim/article/view/316/pdf_94

Conflicto de intereses

Los autores declaran que no existe conflicto de intereses.

Contribuciones de los autores

Nelsa María Sagaró del Campo: Revisión, análisis y selección bibliográfica, revisión y corrección del informe y confección del informe final.

Larisa Zamora Matamoros: Revisión, análisis y selección bibliográfica, revisión y corrección del informe y confección del informe final.