

Estadística médica

Recopilación con fines didácticos



¹Juan O. Talavera,
²Rodolfo Rivas-Ruiz

Investigación clínica IV. Pertinencia de la prueba estadística

¹Centro de Adiestramiento en Investigación Clínica,
 Coordinación de Investigación en Salud,

²Servicio de Escolares y Adolescentes, Hospital de Pediatría

Autores 1 y 2, Centro Médico Nacional Siglo XXI,
 Instituto Mexicano del Seguro Social, Distrito Federal, México

Comunicación con: Juan O. Talavera
 Correo electrónico: jotalaverap@uaemex.mx

Resumen

Cuando observamos la diferencia entre dos terapias o la asociación de un factor de riesgo o indicador pronóstico con su desenlace, tenemos que evaluar la certeza del resultado. Esta evaluación se basa en un juicio que utiliza información relacionada con el diseño del estudio y el manejo estadístico de la información. En este artículo se menciona específicamente la pertinencia de la prueba estadística seleccionada. Las pruebas estadísticas se eligen a partir de dos características: el objetivo del estudio y el tipo de variables. El objetivo se puede dividir en tres grupos de pruebas: *a*) en las que se desea mostrar diferencias entre grupos, o de un mismo grupo antes y después de una maniobra; *b*) en las que se busca mostrar la relación entre las variables; *c*) en las que se pretende predecir un desenlace. En cuanto a los tipos de variables tenemos dos: las cuantitativas (continuas y discontinuas) y las cualitativas (ordinales y dicotómicas). Por ejemplo, si se busca demostrar diferencias en edad (variable cuantitativa), entre pacientes con lupus eritematoso sistémico, con y sin afección neurológica (dos grupos), la prueba apropiada es la *t* de Student para muestras independientes; pero si lo que se compara entre esos mismos grupos es la frecuencia de mujeres (variable binomial), entonces la prueba estadística pertinente es la χ^2 .

Palabras clave

investigación biomédica
 proyectos de investigación
 estadística y datos numéricos

Summary

When we look at the difference between two therapies or the association of a risk factor or prognostic indicator with its outcome, we need to evaluate the accuracy of the result. This assessment is based on a judgment that uses information about the study design and statistical management of the information. This paper specifically mentions the relevance of the statistical test selected. Statistical tests are chosen mainly from two characteristics: the objective of the study and type of variables. The objective can be divided into three test groups: *a*) those in which you want to show differences between groups or inside a group before and after a maneuver, *b*) those that seek to show the relationship (correlation) between variables, and *c*) those that aim to predict an outcome. The types of variables are divided in two: quantitative (continuous and discontinuous) and qualitative (ordinal and dichotomous). For example, if we seek to demonstrate differences in age (quantitative variable) among patients with systemic lupus erythematosus (SLE) with and without neurological disease (two groups), the appropriate test is the "Student *t* test for independent samples". But if the comparison is about the frequency of females (binomial variable), then the appropriate statistical test is the χ^2 .

Key words

biomedical research
 research design
 statistics and numerical data

Introducción

Cuando observamos la diferencia entre dos terapias o la asociación de un factor de riesgo o indicador pronóstico con su desenlace, surge un cuestionamiento: ¿el resultado es real? Decidir si es real requiere dos juicios que se complementan:

1. La planeación y el desarrollo del proceso que documentan dicha diferencia o asociación están libres de errores, o cuando menos éstos son de una magnitud menor que no modifican la orientación de la diferencia o de la asociación (es decir, diseño apropiado y ejecución adecuada).

2. El tamaño de la muestra es suficiente para mantener la estabilidad de los datos y la prueba estadística es apropiada al objetivo.

La planeación y el desarrollo del proceso se han mencionado en los tres capítulos previos de esta serie. Por otra parte, la estabilidad de los datos será discutida a detalle en un artículo posterior al hablar de tamaño de muestra y valor de *p*.

En este artículo discutiremos la pertinencia de la prueba estadística seleccionada. Sin duda, este conocimiento nos permitirá comprender con mayor precisión los resultados obtenidos en estudios de investigación clínica y, por supuesto, incrementará nuestra capacidad para hacer un uso adecuado de los mismos.

Objetivo del estudio y tipo de variable

Las pruebas estadísticas se eligen a partir de dos características: el objetivo del estudio y el tipo de variables. Dentro de los objetivos del estudio podemos identificar tres:

1. Demostrar diferencias entre grupos o diferencias en un mismo grupo antes y después de una maniobra (por ejemplo, el tratamiento con el fármaco *A* reduce en mayor proporción la presión sanguínea que el tratamiento con el fármaco *B*).

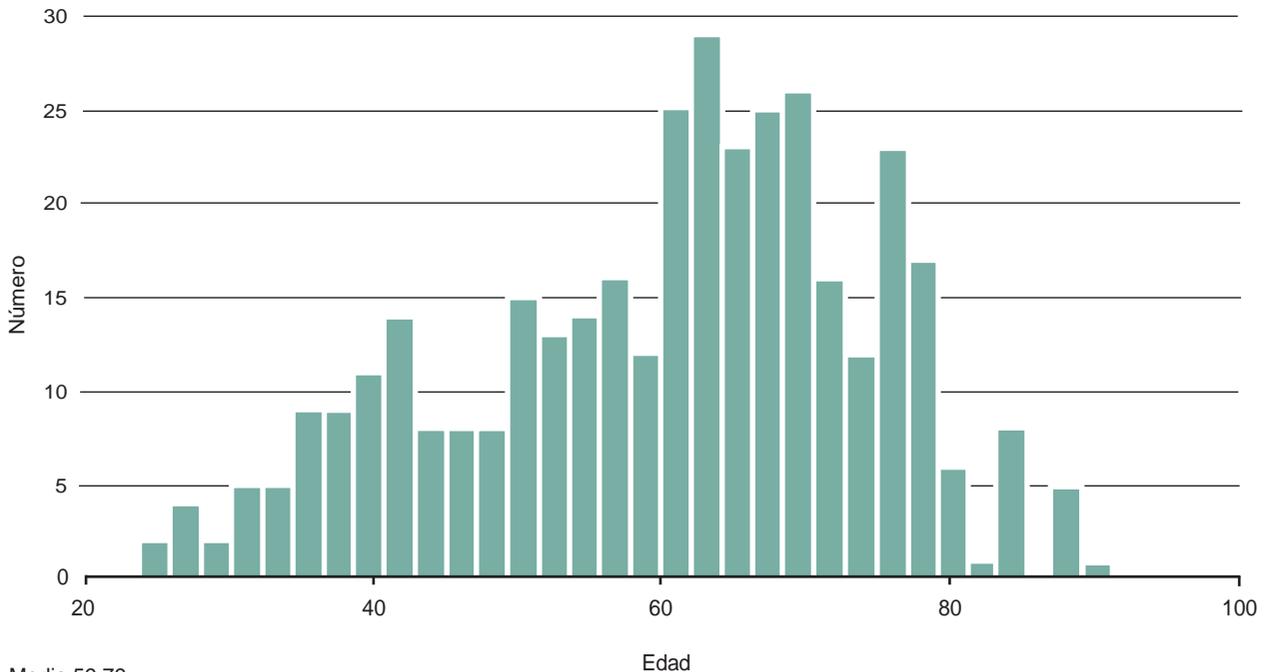
Cuadro I | Peso de sujetos en estudio bajo dos esquemas terapéuticos

Grupo A			Grupo B		
77			65		
78			69		
80			77		
82			78		
85	83.5	Promedio	85	83.5	Promedio
85	85.0	Mediana	85	85.0	Mediana
85	85.0	Moda	89	85.0	Moda
86			93		
88			96		
89			98		

Las medidas de tendencia central son iguales, pero la dispersión de los datos es distinta

- Mostrar relación (correlación) entre variables (por ejemplo, la creatinina sérica se eleva conforme disminuye la función renal).
- Predecir un desenlace (por ejemplo, la probabilidad de que el sujeto con vida sedentaria y sobrepeso desarrolle diabetes mellitus tipo 2).

Frecuentemente se superponen los modelos, y es así que en ocasiones para demostrar diferencias entre dos grupos se hace uso de modelos inicialmente identificados para prede-



Media 59.79
 Desviación estándar de 13.882. Dos desviaciones estándar a cada lado de la media reflejan 95 % de la población
 Promedio de 59.79, con IC 95 % = 32.03-87.55

Figura 1 | Histograma

cir un desenlace; esto ocurre sobre todo cuando se requiere ajustar la maniobra principal (fármaco A versus fármaco B) por múltiples factores (edad, sexo, índice de masa corporal, etcétera). Pero también acontece el fenómeno opuesto, cuando se busca predecir un suceso que ocurrirá a futuro pero solo se cuenta con una o dos variables predictivas, en este caso se utiliza una prueba para demostrar diferencias.

Es importante aclarar que la correlación básicamente sirve para ver la magnitud de la asociación entre variables, aunque debe quedar claro que no determina causalidad, de hecho ninguna prueba estadística puede hacerlo, para ello se requiere cubrir una serie de principios descritos por sir Austin Bradford Hill.

Definir el *tipo de variable* resulta relevante debido a que es el eje para seleccionar la prueba apropiada dependiendo del objetivo buscado. Dentro del tipo de variable existen dos grupos:

1. *Cuantitativas continuas y discontinuas o discretas*: las primeras se caracterizan porque pueden tomar cualquier valor a lo largo de un continuo (por ejemplo, talla 1.75 m). Por su lado, las discontinuas o discretas utilizan solo números enteros (número de partos, 1, 2, 3...). En ambas, la distancia entre una unidad y otra a través de toda su escala es equidistante.
2. *Cualitativas*: dentro de éstas se encuentran la ordinal y la dicotómica. La *ordinal* permite dar un orden a la característica en estudio, y a diferencia de las cuantitativas la distancia entre dos categorías no es equidistante (por ejemplo, falla cardíaca grados I a IV). Las *dicotómicas*, como su nombre lo dice, son en las que solo hay dos categorías, que pueden ser binomiales (es una opción u otra, por ejemplo, hombre o mujer) o nominales (se refiere a la presencia o no de la característica, por ejemplo, vivo a seis meses, sí o no).

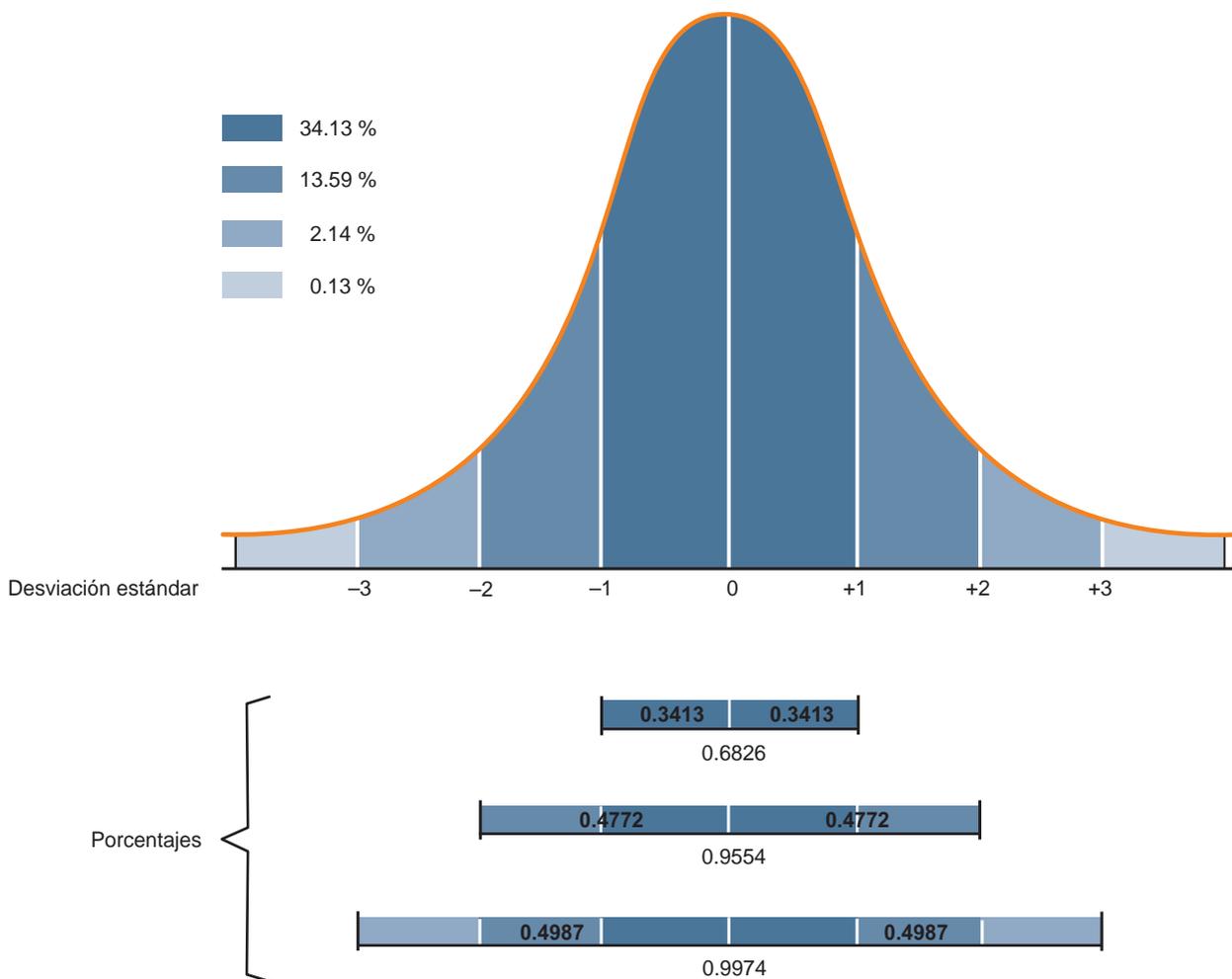


Figura 2 | Curva de distribución normal

Cuadro II | Selección de la prueba estadística conforme el objetivo y el tipo de variable

Tipo variable	Tipo muestra	Demostrar diferencia		Mostrar relación ^{&}	Predecir 1 variable [‡]
		Dos grupos	Tres grupos	Dos variables	Variable desenlace
Cuantitativa (distribución normal)	NR	t Student*	Anova 1 factor	Pearson	Regresión lineal
	R	t Student**	Anova 1 factor		
Cualitativa ordinal	NR	U Mann-Whitney	Kruskal-Wallis	Spearman	
	R	Wilcoxon	Friedman		
Cualitativa dicotómica R	NR	χ^2 (Fisher)	χ^2	Coeficiente phi Curvas de supervivencia	Regresión logística/
	R	McNemar			

NR = no relacionada, R = relacionada, R = medición de la variable en el mismo sujeto en dos tiempos distintos
 *t de Student para muestras independientes
 **t de Student para muestras relacionadas
[&]Para la correlación entre 2 variables se utiliza la prueba de aquella en la escala inferior (en realidad ninguna escala es inferior, sin embargo, se han ordenado de las cuantitativas continuas a las dicotómicas, pasando por las cuantitativas discontinuas y las ordinales).
[‡]La variable predictiva puede ser cuantitativa, dicotómica u ordinal (estas últimas transformadas a variables tipo dummy)

Es importante mencionar el manejo que sufrirá el tipo de variable durante el proceso de análisis, empezando por la recolección de datos en forma “cruda”, lo que significa que solo se trata de una compilación de información de un grupo de sujetos. Para que estos datos tengan un significado de utilidad se necesita organizarlos y resumirlos. La forma más simple de organización son las tablas de distribución de frecuencia, sin embargo, en ocasiones es más fácil entender su representación gráfica a través de un histograma o polígono de frecuencias. No obstante la utilidad de esta información, se necesita información cuantitativa de los datos recolectados, es decir, se requieren *índices numéricos* que reflejen las distintas distribuciones de probabilidad, cuya función primordial es *modelar* el comportamiento de una gran variedad de fenómenos biológicos. Dentro de estos índices numéricos encontramos las medidas de tendencia central y las medidas de dispersión.

1. Medidas de tendencia central (cuadro I y figura 1).

- a) *Media*: es la suma de un conjunto de datos divididos por el número total de ellos. El símbolo para representar la media de una población es la letra griega mu (μ), y la media de una muestra está representada por \bar{X} . Es la medida de resumen más utilizada para variables cuantitativas.
- b) *Mediana*: es el valor que se encuentra exactamente a la mitad del total de los datos. La mediana divide una distribución de datos ordenados en exactamente dos partes iguales. La ventaja de la mediana como medida de tendencia central es que ésta no se ve afectada por el valor de datos extremos, fenómeno que sí ocurre con la media. Es el tipo de medida de resumen más utilizado para

las variables cuantitativas que no siguen una distribución normal y para variables ordinales.

- c) *Moda*: se refiere al valor que más se repite en una distribución.

2. Medidas de dispersión más comunes.

- a) *Desviación estándar*: refleja la variación entre todo el conjunto de datos y se utiliza cuando éstos siguen una distribución normal.
- b) *Percentil*: describe la posición de un valor de la distribución. Se utiliza para variables cuantitativas que no siguen una distribución normal y para variables de tipo ordinal.
- c) *Rango*: diferencia entre el valor más alto y el valor más bajo de la distribución.

En investigación clínica, como en muchos otros fenómenos de la vida real, los datos más frecuentemente analizados son los cuantitativos, que en el más de los casos muestran una distribución gaussiana, también conocida como distribución normal, que se caracteriza por tener forma acampanada, ser simétrica respecto a su media, tener valores de frecuencia decrecientes conforme se alejan de la media y que nunca llegan a cero (asintótica), la moda y la mediana son iguales a la media, aproximadamente 68 % de los datos se encuentran en ± 1 desviación estándar respecto a su media y 95 % en ± 2 desviaciones estándar (figura 2). Es así que si la agrupación de datos son de tipo cuantitativo con distribución normal, su medida de resumen será la media y su medida de dispersión, la desviación estándar. Sin embargo, si su distribución no es gaussiana, al igual que para una variable de tipo ordinal, su

medida de resumen será la mediana y su medida de dispersión, el percentil o rango. Finalmente, habrá que mencionar que para las variables dicotómicas su medida de resumen es la frecuencia y su medida de dispersión la desviación estándar (su distribución se acerca a la normal).

Pertinencia de la prueba estadística

Una vez que conocemos nuestro objetivo y las características de nuestros datos (tipo de variable) podemos considerar la pertinencia de la prueba estadística (cuadro II). Sin embargo, existen dos consideraciones más cuando el objetivo sea demostrar diferencia:

1. Si se trata de un estudio donde se compara el valor de un dato antes y después de una maniobra ya sea observacional o experimental, se le conoce como de muestras relacionadas, pero si se trata de la comparación de datos entre distintos grupos se le denomina de muestras no relacionadas.
2. Si se trata de comparación entre distintos grupos es necesario determinar si será entre dos o más grupos.

Ya con la información completa, con el cuadro II podemos verificar si la selección de la prueba estadística fue la adecuado conforme la variable y el objetivo. Por ejemplo, si se compara la edad (variable cuantitativa con distribución normal en este caso), entre pacientes con lupus eritematoso sistémico, con y sin afección neurológica (dos grupos), la

prueba apropiada es la *t* de Student para muestras independientes. Pero si lo que se compara entre esos mismos es la frecuencia de mujeres (variable binomial), entonces la prueba estadística pertinente es la χ^2 . Si lo que se compara entre los dos grupos es su grado de actividad lúpica (escala ordinal), la prueba estadística pertinente es la *U* de Mann-Whitney. Por otro lado, si lo que nos muestran es la magnitud de la asociación (relación) entre edad (variable cuantitativa con distribución normal) y el grado de actividad lúpica (variable ordinal), la prueba pertinente es la *r* de Spearman. Finalmente, si lo que se busca es predecir el peso de un niño (variable cuantitativa) a partir de la edad (variable cuantitativa), tipo de alimentación (variable ordinal: buena, regular o mala) y el sexo (dicotómica), la prueba pertinente es la regresión lineal. Pero si lo que se quiere es predecir la probabilidad de infarto (dicotómica nominal) en los siguientes 10 años a partir de la edad (cuantitativa), el riesgo aterogénico (ordinal, bajo, medio y alto) y el sexo (dicotómica binomial), la prueba pertinente es la regresión logística múltiple.

Por último, esperamos que el presente artículo permita entender la razón de la selección de las pruebas estadísticas más empleadas en la investigación en salud y que, de igual manera, sirva de guía a quienes se inician en la estadística. No es suficiente para definir si los resultados obtenidos son reales, será necesario tener en consideración el diseño y ejecución del estudio y la estabilidad de los datos, pero este último aspecto merece ser analizado en otro apartado.

Bibliografía

1. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. 4th ed. Massachusetts, MA: Blackwell Publishing; 2002.
2. Bland M. Introduction to medical statistics. 3rd ed. Oxford: Oxford University Press; 2003.
3. Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia, PA: W.B. Saunders; 1985.
4. Feinstein AR. Multivariable analysis: an introduction. New Haven, CT: Yale University Press; 1996.
5. Feinstein AR. Principles of medical statistics. New York, NY: Chapman and Hall/CRC; 2002.
6. Le Chap T. Introductory biostatistics. Hoboken, NJ: New Jersey: John Wiley and Sons; 2003.
7. Peat J, Barton B. Medical statistics. A guide to data analysis and critical appraisal. Malden, MA: Blackwell Publishing; 2005.
8. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. 3rd ed. Saddle River, NJ: Pearson/Prentice Hall; 2009.



Investigación clínica XIV Del juicio clínico al modelo estadístico

Juan O. Talavera,^a Rodolfo Rivas-Ruiz^a

Clinical research XIV. From the clinical judgment to the statistical model

A statistical test is incomprehensible when it is out of context, so it is necessary to identify the details of the phenomenon of causality in the clinical course of the disease and to integrate the statistical model. Thus, the statistical tests used will try to characterize baseline, maneuver and the outcome, and will show the relationship between them. When we read the results in clinical research, the first thing that the author describes are general characteristics of the population, starting with number of patients evaluated and selected, average age, gender, and number of subjects meeting the outcome. It is extremely important because with the same criteria two studies may contain populations completely opposite. Posterior description usually continues through tables that follow a logical sequence, which allows us to integrate the statistical model to clinical judgment: baseline characteristics of the population and its distribution in each of the maneuvers, characteristics of the main and peripheral maneuvers, main effect of the maneuver on the outcome, and the impact of principal maneuver in the outcome, but adjusted for any variable that can alter this impact.

Key words

biomedical research
research design
statistics as topic

En la parte IV de esta serie, “Pertinencia de la prueba estadística”, se mencionó la razón de ser de la prueba estadística y los tres objetivos generales que se pretende alcanzar con ella: mostrar diferencias, mostrar asociación y predecir un fenómeno. Las pruebas orientadas al último objetivo (por ejemplo, los modelos de regresión lineal, los de regresión logística y las curvas de supervivencia) permiten mostrar diferencias entre los grupos al hacer los ajustes conforme las distintas variables involucradas.

Si bien es posible identificar qué quiere decir una prueba estadística, al estar fuera de contexto pierde su significado. De tal suerte, antes debe considerarse la pregunta de investigación y darle un contexto al análisis estadístico dentro del juicio clínico. Para ello es necesario reflexionar nuevamente de qué forma en los estudios de causalidad se busca demostrar cómo una maniobra (que puede ser observacional o experimental) provoca un desenlace al intervenir sobre el estado basal (figura 1).

Del juicio clínico al modelo estadístico

En la descripción de los resultados de una investigación, en primer término se describen las características generales de la población en estudio, comenzando por el número de pacientes evaluados y las razones de la selección, la edad promedio de la población, el sexo y el número de sujetos que sufrieron el desenlace. Este apartado es relevante ya que dos estudios pueden contener en su interior poblaciones completamente opuestas aun cuando se llevaron a cabo bajo los mismos criterios.

Posterior a la descripción inicial, para la exposición de la información se recurre a cuadros que siguen una secuencia lógica y permiten claramente integrar el modelo estadístico al juicio clínico:

- En el primer cuadro se describen las características basales de la población y, opcionalmente, cómo quedaron distribuidas en cada una de las maniobras (trátase de una maniobra observacional o experimental).
- En el segundo cuadro se muestran las características de las maniobras principal (apego o efecto sobre los reguladores intermedios) y periférica.
- El tercer cuadro se destina al efecto de la maniobra principal sobre los desenlaces primario y secundario (en ocasiones se realiza un cuadro para cada uno de ellos, sobre todo cuando en los desenlaces secundarios influyen los eventos adversos de forma importante).
- En el cuarto cuadro se describe el impacto de la maniobra principal en el desenlace, pero ajustando

Una prueba estadística fuera de contexto resulta incomprensible, por ello es necesario identificar los pormenores del fenómeno de causalidad dentro del curso clínico de la enfermedad e integrar a este el modelo estadístico. De tal forma, con las pruebas estadísticas se buscará caracterizar el estado basal, la maniobra y el desenlace, y mostrar la relación entre ellos. En la exposición de los resultados de una investigación, en primer término se deben describir las características generales de la población, comenzando por el número de pacientes evaluados y seleccionados, la edad promedio, el sexo y el número de sujetos que presentaron el desenlace. Lo anterior es relevante ya que dos estudios pueden contener poblaciones completamente opuestas aun cuando se realicen bajo los

mismos criterios. Generalmente, la descripción de los resultados se continúa mediante cuadros, los cuales siguen una secuencia lógica y permiten integrar claramente el modelo estadístico al juicio clínico: características basales de la población y su distribución en cada una de las maniobras, características de las maniobras (principal y periféricas), efecto de la maniobra principal sobre el desenlace y su impacto ajustado conforme todas las variables que puedan modificar dicho impacto

Resumen

Palabras clave

investigación biomédica
proyectos de investigación
estadística como asunto

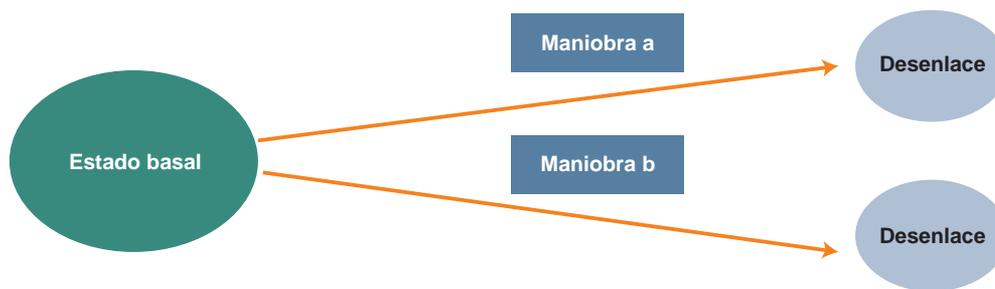


Figura 1 Componentes del estudio de causalidad

el resultado conforme a todas las variables que puedan modificar dicho impacto, es decir, a las características presentes en el estado basal y a las maniobras periféricas.

Evidentemente hay variaciones en esta estructura base (este orden se sigue fielmente en los artículos publicados en el *New England Journal of Medicine*, por ejemplo en el estudio JUPITER [N Engl J Med. 2008;359:2195-207]) y se pueden fragmentar o combinar cuadros o combinar, sustituir o complementar con gráficos. Para efecto de ejemplificar, en el desarrollo de esta exposición se utilizará el ensayo clínico “Reduction in the incidence of poststroke nosocomial pneumonia by using the ‘turn-mob’ program” (J Stroke Cerebrovasc Dis. 2010;19(1):23-8), utilizado previamente en esta serie.

Como se mencionó, en el primer cuadro pueden describirse las características basales de la población y la distribución de estas de acuerdo con la maniobra principal —con el fin de mostrar que dichas características están distribuidas homogéneamente entre los grupos—. Con frecuencia, el orden de la presentación de los datos sigue la misma disposición en la que se obtiene la información durante la atención médica: datos generales, síntomas y antecedentes, signos clínicos y datos de laboratorio y gabinete

(figura 2). En el ejemplo del programa *turn-mob* se describen y comparan las características basales entre las maniobras, empezando por los datos generales, los antecedentes, los signos y los resultados de los exámenes de gabinete; no hay una evaluación de síntomas debido a la alteración neurológica de los pacientes.

En el segundo cuadro se describe la calidad de las maniobras principal (apego) y periféricas (figura 3). En el ejemplo no se señala el grado de apego a la maniobra principal ni a la comparativa como una omisión del artículo, pero se muestra la distribución de las maniobras periféricas entre los grupos.

En el tercer cuadro se evidencia el efecto de la maniobra principal sobre los desenlaces primario y secundario sin hacer ajustes por las posibles variables de confusión (figura 4). Debe quedar claro que la evaluación del desenlace se realizó de igual forma en ambos grupos (si bien la estrategia de evaluación se desglosa en el apartado relativo a los métodos, en el apartado de los resultados se deberá documentar en cuántos pacientes se logró tal objetivo) y mencionar cuántos pacientes se perdieron en el seguimiento; para estos se deben comparar las características presentes hasta el último seguimiento entre los pacientes que se continuó estudiando y aquellos en los que no fue posible continuar el análisis. Estos dos últimos aspectos se pueden mencionar en el texto de la investigación,

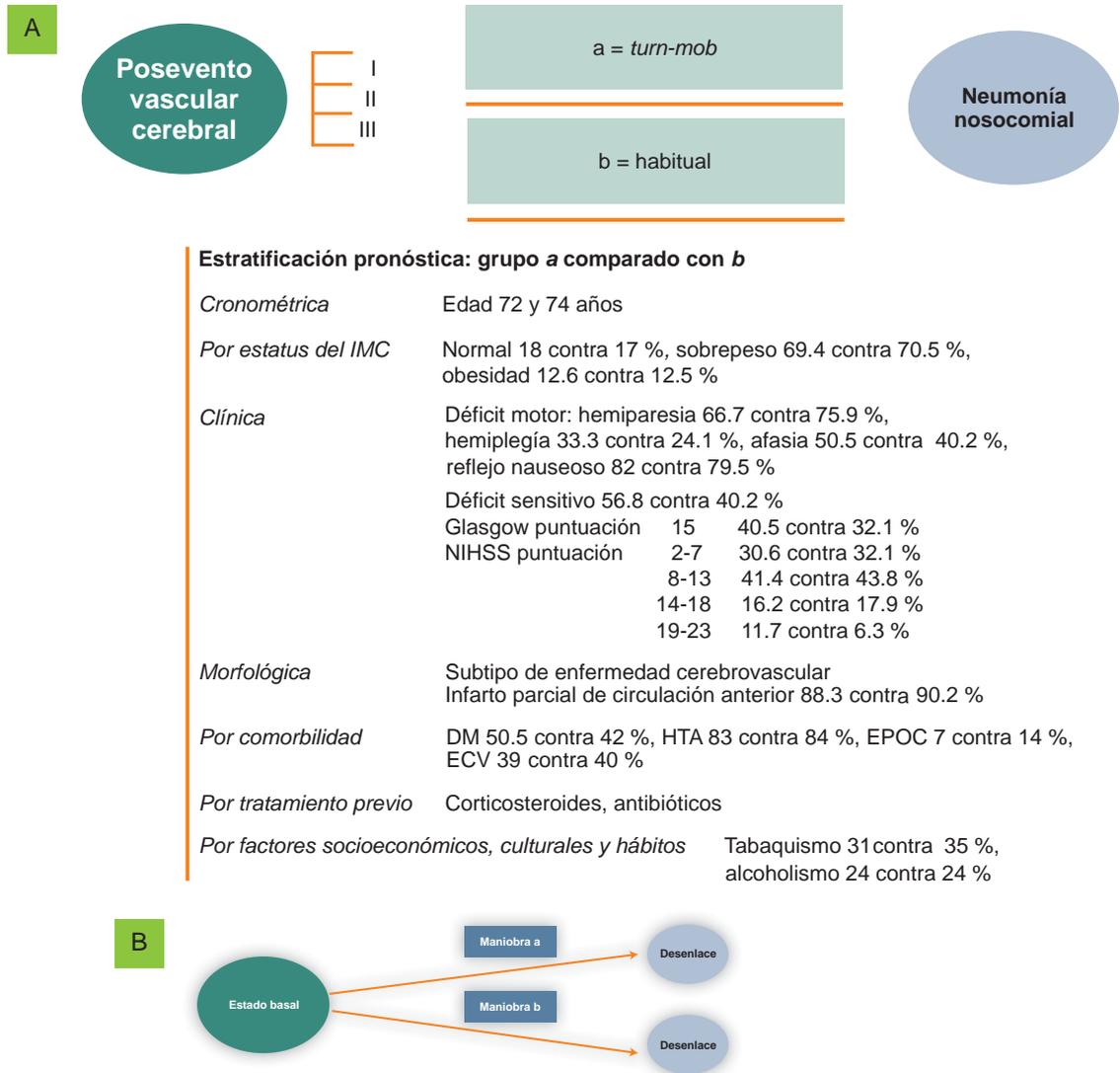


Figura 2 A) Ejemplo de las características basales en el artículo acerca del programa *turn-mob*. B) Fragmento del cuadro de las características basales (publicado en el mismo artículo)

al final del párrafo que describe el cuadro correspondiente o en el pie del cuadro.

Frecuentemente se suele estructurar un cuadro para el desenlace secundario cuando este se refiere a eventos adversos y la variedad y frecuencia de estos son elevadas. En el ejemplo se muestra el efecto de la maniobra principal sobre el desenlace sin realizar ningún ajuste. En el texto se menciona que se excluyeron dos pacientes por presentar neumonía durante las pri-

meras 48 horas, lo que los excluye del diagnóstico de neumonía nosocomial

Finalmente, en el cuarto cuadro se desglosa el impacto de la maniobra principal en el desenlace ajustando los resultados a las características presentes en el estado basal y a las maniobras periféricas que puedan tener impacto (figura 5). Por lo general se requiere el análisis multivariado para este ajuste y se deben incluir todas las demás variables, lo cual puede efectuarse por

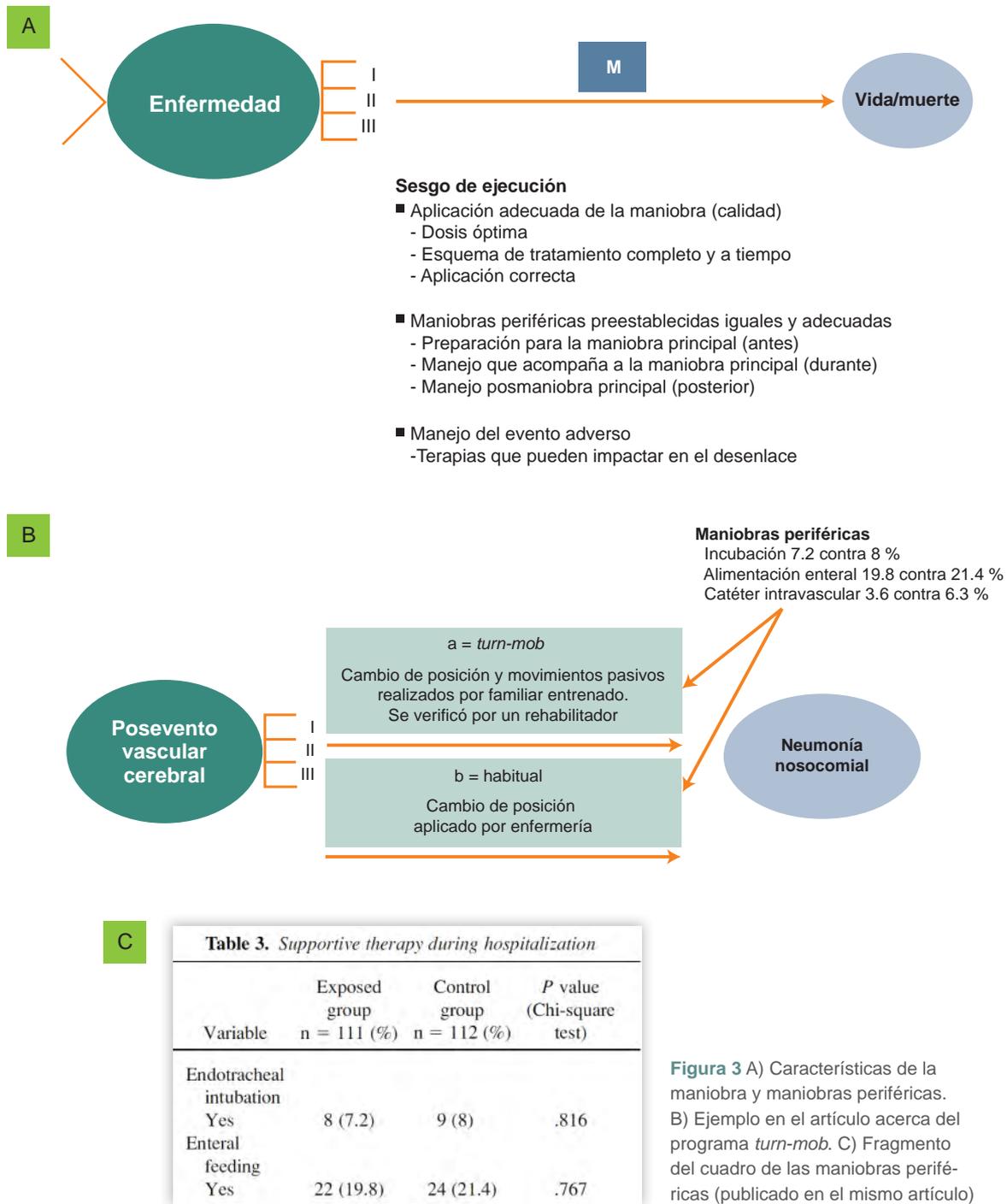
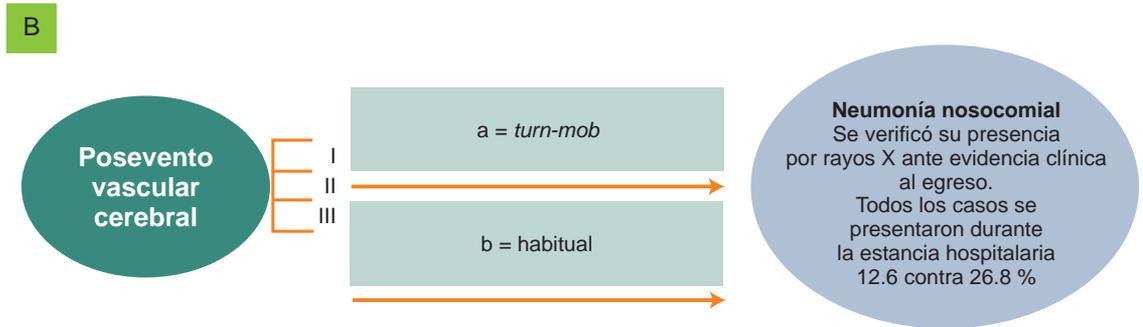
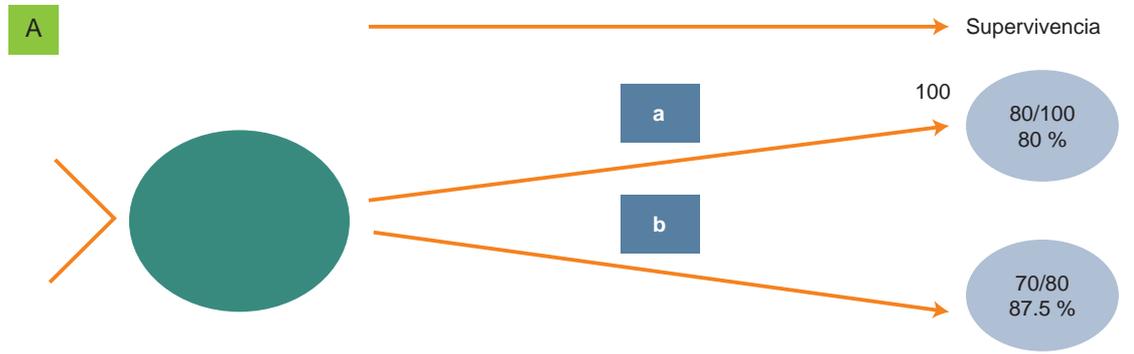


Figura 3 A) Características de la maniobra y maniobras periféricas. B) Ejemplo en el artículo acerca del programa *turn-mob*. C) Fragmento del cuadro de las maniobras periféricas (publicado en el mismo artículo)

pasos (modelo 1, solo maniobra; modelo 2, maniobra y características basales; modelo 3, maniobra, características basales y maniobra periférica, etcétera). La finalidad es mostrar el efecto modificador de las características específicas que se van agregando durante el curso clínico de la enfermedad.

En ocasiones se realiza un ajuste por cada una de estas características y gráficamente se muestra como si se efectuara con todas a la vez, sugiriendo falsamente

un ajuste completo. En otras ocasiones no se realiza este ajuste debido a que los autores muestran que las características basales y las maniobras periféricas quedaron bien balanceadas entre los grupos en comparación. Cuando sucede así, como en el programa *turn-mob*, es probable que se pierda la oportunidad de observar cómo algunas características pueden modificar el impacto de la maniobra y su efecto sobre el desenlace.



C

Table 4. Contingency table of maneuver-associated nosocomial pneumonia

	Pneumonia		Totals
	Yes	No	
Exposed	14	97	111
Control	30	82	112

Figura 4 A) Efecto de la maniobra principal sobre el desenlace. B) Ejemplo en el artículo acerca del programa *turn-mob*. C) Fragmento del cuadro acerca del impacto de la maniobra principal sobre el desenlace (publicado en el mismo artículo)

Comentarios

Si bien el presente texto trata de explicar cómo en un artículo se integra la estadística al razonamiento clínico (juicio clínico, arquitectura de la investigación), la elección de la prueba estadística depende de los criterios específicos mencionados en capítulos anteriores de esta serie (partes IV, V y VI), en los que se ahondará posteriormente. Por fortuna para el médico, esta integración resulta lógica y fácil de entender y confirma por qué son insu-

ficientes las respuestas que solo mencionan el beneficio de la maniobra principal y soslayan el efecto de todas las condiciones del paciente, de la enfermedad y de las terapias periféricas, que no en raras ocasiones explican el desenlace.

Declaración de conflicto de interés: los autores han completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no fue reportado alguno en relación con este artículo.

^aCentro de Adiestramiento en Investigación Clínica, Coordinación de Investigación en Salud, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Distrito Federal, México

Comunicación con: Juan O. Talavera
Correo electrónico: jotalaverap@uaemex.mx

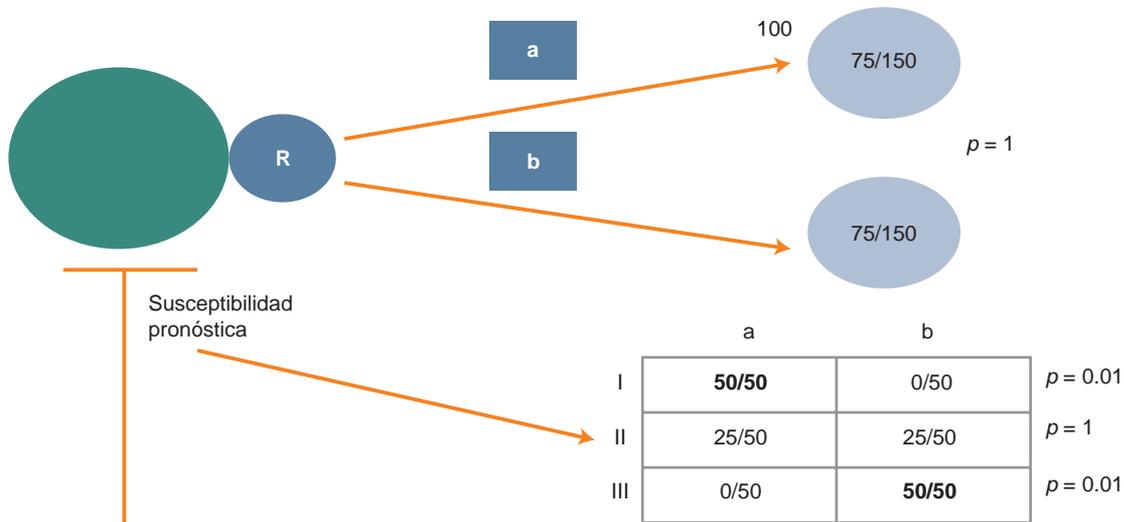


Figura 5 Efecto de la maniobra principal sobre el desenlace, ajustado conforme las características basales y maniobras periféricas

Lecturas recomendadas

1. Feinstein AR. Clinical biostatistic. Washington: Mosby; 1977.
2. Feinstein AR. Clinical epidemiology. The architecture of clinical research. Philadelphia: WB Saunders; 1985.
3. Feinstein AR. Clinical judgment. Baltimore: Williams and Wilkins; 1967.
4. Feinstein AR. Multivariable Analysis: an introduction. New Haven, CT: Yale University Press; 1996.
5. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Third edition. New Jersey: Pearson/Prentice Hall; 2009.
6. Talavera JO. Clinical research I. The importance of the research design. Rev Med Inst Mex Seguro Soc. 2011;49(1):53-8.
7. Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. Rev Med Inst Mex Seguro Soc. 2011;49(3):289-94.
8. Talavera JO, Rivas-Ruiz R. Clinical research IV. Relevancy of the statistical test chosen. Rev Med Inst Mex Seguro Soc. 2011;49(4):401-5
9. Talavera JO, Rivas-Ruiz R, Bernal-Rosales LP. Clinical research V. Sample size. Rev Med Inst Mex Seguro Soc. 2011;49(5):517-22
10. Talavera JO, Rivas-Ruiz R. Clinical research VI. Clinical relevance. Rev Med Inst Mex Seguro Soc. 2011;49(6):631-5.



Investigación clínica XV Del juicio clínico al modelo estadístico. Diferencia de medias. Prueba *t* de Student

Rodolfo Rivas-Ruiz,^a Marcela Pérez-Rodríguez,^a Juan O. Talavera^a

Clinical research XV. From the clinical judgment to the statistical model. Difference between means. Student's *t* test

Among the test to show differences between means, the Student *t* test is the most characteristic. Its basic algebraic structure shows the difference between two means weighted by their dispersion. In this way, you can estimate the *p* value and the 95 % confidence interval of the mean difference. An essential feature is that the variable from which the mean is going to be calculated must have a normal distribution. The Student *t* test is used to compare two unrelated means (compared between two maneuvers), this is known as *t* test for independent samples. It is also used to compare two related means (a comparison before and after a maneuver in just one group), what is called paired *t* test. When the comparison is between more than two means (three or more dependent means, or three or more independent means) an ANOVA test (or an analysis of variance) it is used to perform the analysis

Key words

T test
normal distribution
statistics
ANOVA

Una vez que hemos entendido cómo es que se genera la propuesta de análisis estadístico (modelo estadístico) y el tipo de prueba estadística que se va a utilizar, dependiendo del objetivo de la prueba y del tipo de variable, vale la pena entender con mayor detalle qué hay al interior de las pruebas estadísticas utilizadas con mayor frecuencia en investigación clínica.

Ya en “Investigación clínica IV” de esta serie se discutió sobre la prueba estadística pertinente de acuerdo con el tipo de variable y su objetivo. Dentro de este artículo se identificaron tres grupos distintos: 1) cuando se busca mostrar diferencia entre grupos, 2) la relación entre dos variables, o 3) cuando se busca predecir un desenlace.

En ese primer grupo, existen tres tipos de pruebas que dependen de la variable y de la distribución de estas.

La primera corresponde a las pruebas cuya variable es cuantitativa y, además, se presenta con distribución normal, es decir, tiene forma acampanada, es simétrica con respecto a la media, tiene valores de frecuencia decreciente conforme se aleja de esta y nunca llega a cero (es asintótica). Asimismo, la moda y la mediana son iguales a la media. Dentro de una desviación estándar se encuentra el 68 % de los datos y dentro de dos desviaciones estándar el 95 % (ver figura 2, p. 403, en “Investigación clínica IV. Pertinencia de la prueba estadística”, publicada en *Rev Med Inst Mex Seguro Soc* 2011;49(4):401-405).

Diferencia de medias. Prueba *t* de Student

Entre las pruebas para demostrar diferencia de medias, la más característica es la prueba *t* de Student para muestras independientes. Imaginemos como ejemplo una población base de pacientes hipertensos con cifras promedio de presión diastólica de 95 mm Hg. Esta población es dividida al azar en dos grupos de tratamiento: antihipertensivo 1 (AH1) y antihipertensivo 2 (AH2). Después de tres meses de tratamiento, se observa que el grupo que recibió el AH1 tiene una presión diastólica de 85 mm Hg, mientras el que recibió el AH2, de 94 mm Hg, lo que hace una diferencia de medias entre los dos grupos de -9 mm Hg. Si bien resulta clínicamente relevante y parece que ahora se trata de dos poblaciones distintas, no sabemos si la diferencia es al azar y en realidad ambos grupos continúan siendo una misma población. Aquí es donde la prueba *t* de Student para muestras independientes nos permitirá saber si la diferencia es real o es al azar (figuras 1a y 1b). El valor de $p \leq 0.001$ nos muestra que la diferencia es real y en consecuencia se puede asumir diferencia entre los tratamientos, lo cual da origen a dos poblaciones distintas.

Dentro de las pruebas para demostrar diferencia de medias, la más característica es la *t* de Student. La estructura algebraica base de esta prueba muestra la diferencia ponderada del promedio de una variable menos el promedio de otra entre su dispersión; de esta manera, se puede calcular el valor de *p* y el intervalo de confianza de 95 % para dicha diferencia de medias. Una característica indispensable es que la variable de la cual se va a calcular la media tenga distribución normal. La prueba *t* de Student igual se utiliza para dos medias de muestras no relacionadas (se compara entre dos maniobras) a lo que se le conoce como prueba *t* para muestras independientes, o para

dos medias de muestras relacionadas (una comparación de antes y después de una maniobra), a lo que se le denomina *t* pareada. Cuando la comparación va más allá de dos medias (tres medias dependientes, o tres medias de grupos distintos) el análisis a realizar es un ANOVA (analysis of variance, por sus siglas en inglés).

Palabras clave

Prueba *t*
distribución normal
estadística
ANOVA

Resumen

En el ejemplo de “*Reduction in the incidence of poststroke nosocomial pneumonia by using the ‘Turn-Mob’ program*”, publicado en *Journal of Stroke and Cerebrovascular Disease* 2010;19:23-28, el cual hemos utilizado como ejemplo en esta serie, la edad es una variable cuantitativa con distribución normal y se compara entre las distintas maniobras; en ese ejemplo por edad parecen corresponder a una misma población (figura 2). Es importante, recordar que el objetivo puede ser como en este caso: mostrar que las características basales son iguales entre ambos tratamientos, para lo cual se esperaba que el valor de la *p* fuese > 0.05, tal como ocurrió.

La estructura algebraica base de la prueba *t* de Student muestra la diferencia entre las medias ponderadas por su dispersión:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{x}_1 - \bar{x}_2}}$$

Como se puede ver, en la fórmula se considera la diferencia de medias del primer grupo (*X*₁) menos el segundo grupo (*X*₂), y se divide entre la desviación estándar ponderada de los dos grupos, a lo que se le ha llamado *error estandarizado de la diferencia de medias*. Asimismo, para una diferencia de medias

se puede calcular su intervalo de confianza de 95 % (IC 95 %); en el ejemplo anterior, para una diferencia de -9 mm Hg (85 ± 1.3 frente a 94 ± 1.3), su IC al 95 % sería de (-10.10 a -7.89).

Ejemplo de la prueba *t* de Student en un programa estadístico electrónico

Veamos un ejemplo con los pasos necesarios en un programa estadístico. Supongamos que se realizó un ensayo clínico aleatorio para pérdida de peso en el que un grupo fue asignado a dieta baja en grasas y el otro a una dieta baja en carbohidratos. La hipótesis es que los participantes que estén en el grupo con dieta baja en carbohidratos tendrán un índice de masa corporal (IMC) menor al finalizar el estudio. Asumiendo que el IMC tuvo una distribución normal, realizamos una prueba de hipótesis *t* de Student para comparación de medias de muestras independientes. Los pasos a seguir en el programa SPSS se muestran en la figura 3.

La media de IMC en el grupo con dieta baja en grasa fue de 37.6 ± 1.6 y en el grupo de dieta baja en carbohidratos fue de 36.1 ± 1.7. La diferencia de medias entre los grupos fue de 1.4 unidades de IMC. Antes de elegir el valor de *t* que vamos a utilizar, es necesario saber si

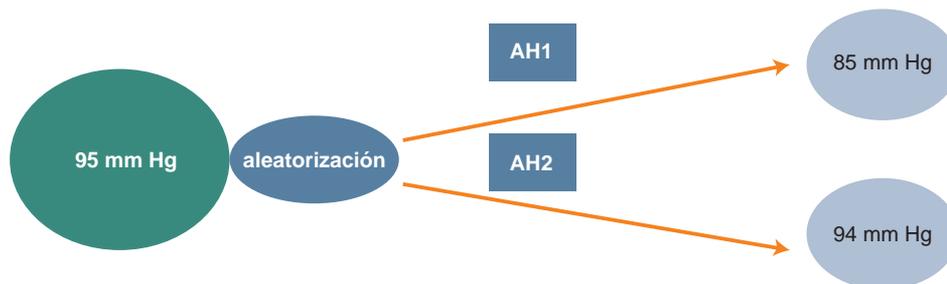


Figura 1a Población base que recibe dos tratamientos asignados al azar (aleatorización) y tres meses después, se evalúa el desenlace

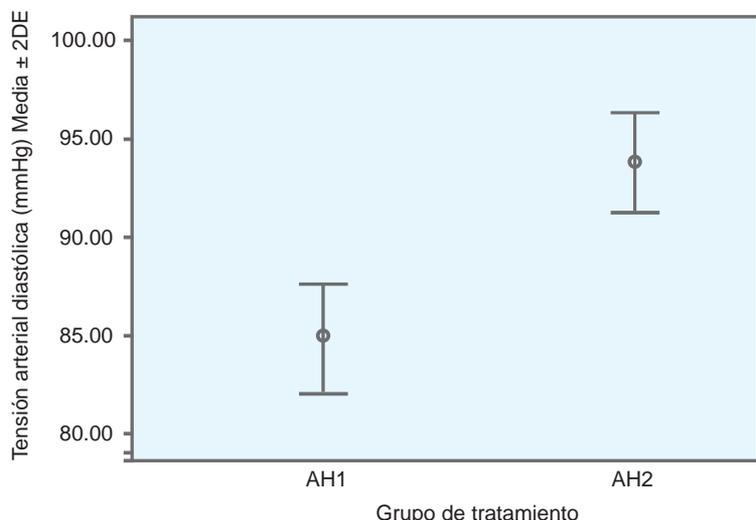


Figura 1b Resultado a tres meses de tratamiento. Prueba t de Student para muestras independientes, $p = 0.000$

Table 1. Baseline characteristics of the study population

Variable	Exposed group n = 111 (%)	Control group n = 112 (%)	P value (Chi-square test)
Age (y)	72.49 ± 10.07*	74.08 ± 9.45*	.224†

Figura 2 Ejemplo del artículo *Turn-mob*. Se muestra que la edad no es diferente entre los grupos de tratamiento

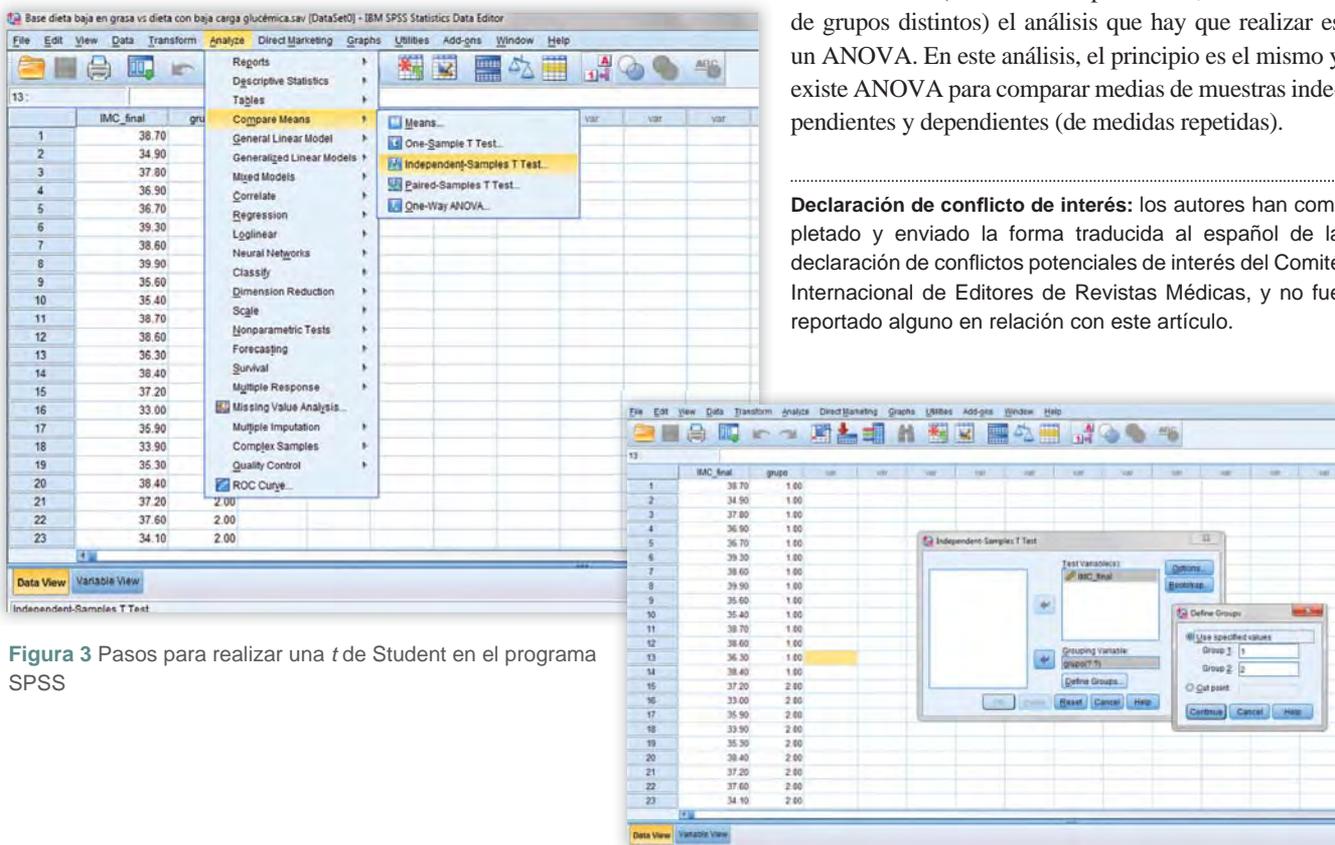


Figura 3 Pasos para realizar una t de Student en el programa SPSS

las varianzas entre los grupos son iguales o diferentes por medio de la prueba de F , que el programa calcula automáticamente (figura 4). En este ejemplo, la significación de la prueba de F es 0.822, lo que quiere decir que las varianzas no son diferentes, ya que necesitaríamos una $p < 0.05$ para asumir que hay diferencias en las varianzas. Dado lo anterior, podemos observar que si asumimos varianzas iguales, la prueba de t nos dio un valor de 2.33 con una $p = 0.028$, lo que implica que la diferencia de medias entre los grupos es real (son estadísticamente significativas). La pregunta más importante es si estas diferencias son clínicamente relevantes. Para ello, el lector del supuesto estudio tendría que hacer uso de su juicio: ¿tener un IMC de 36.1 en comparación con un IMC de 37.6 tiene relevancia clínica? El grupo que presentó el mayor beneficio aún se encuentra clasificado como “obesidad tipo II”.

Comentarios

Como se puede ver, la prueba t de Student nos permite comparar dos medias con distribución normal, ya sean relacionadas o de muestras independientes. También hay que considerar la igualdad de las varianzas, ya que dependiendo de esta la prueba t de Student tiene expresiones distintas. La igualdad de las varianzas se conoce como *homocedasticidad* y la desigualdad como *heterocedasticidad*. Finalmente, si comparamos más allá de dos medias (tres medias dependientes, o tres medias de grupos distintos) el análisis que hay que realizar es un ANOVA. En este análisis, el principio es el mismo y existe ANOVA para comparar medias de muestras independientes y dependientes (de medidas repetidas).

Declaración de conflicto de interés: los autores han completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no fue reportado alguno en relación con este artículo.

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
IMC_final	Equal variances assumed	.052	.822	2.332	26	.028	1.44286	.61863	.17125	2.71446
	Equal variances not assumed			2.332	25.872	.028	1.44286	.61863	.17094	2.71477

Figura 4 Resultado de la prueba t de Student

^aCentro de Adiestramiento en Investigación Clínica, Coordinación de Investigación en Salud, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Distrito Federal, México

Comunicación con: Rodolfo Rivas-Ruiz
Correo electrónico: rivasrodolfo@gmail.com

Referencias

1. Alvan R. Feinstein. Clinical Epidemiology. The architecture of clinical research. Philadelphia, W.B. Saunders Company, 1985.
2. Alvan R. Feinstein. Multivariable Analysis: An introduction. New Haven, Yale University Press, 1996.
3. Portney LG, and Watkins MP. Foundations of clinical research: applications to practice (3er Edition). Pearson/Prentice Hall, 2009.
4. Armitage P, Berry G, and Matthews JNS. Statistical methods in medical research (4th Edition). Massachusetts, Blackwell Publishing Ltd, 2002.
5. Peat J, Barton B. Medical statistics. A guide to data analysis and critical appraisal. Massachusetts, Blackwell Publishing Ltd, 2005.
6. Le Chap T. Introductory biostatistics. New Jersey, John Wiley & Sons, Inc., 2003.
7. Bland M. Introduction to medical statistics (3er Edition). Oxford University Press, 2003.
8. Dawson Beth and Trapp Robert. Basic and clinical biostatistics. 4th edition. Lange Medical Books/McGrawHill. 2004.
9. Portney LG, and Watkins MP. Foundations of clinical research. Applications to practice. Appleton and Lance. 1993.
10. Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. Rev Med Inst Mex Seguro Soc. 2011;49(3):289-94.
11. Feinstein AR. Clinical biostatistic. Washington: C.V. Mosby; 1977.
12. Talavera JO, Rivas-Ruiz R. Clinical research IV. Relevancy of the statistical test chosen. Rev Med Inst Mex Seguro Soc. 2011 Jul-Aug; 49(4):401-5.
13. Talavera JO, Rivas-Ruiz R. Clinical research V. Sample size. Rev Med Inst Mex Seguro Soc. 2011 Sept-Oct;49(5):517-522



Investigación clínica XVI

Diferencias de medianas con la *U* de Mann-Whitney

Rodolfo Rivas-Ruiz,^a Jorge Moreno-Palacios,^a Juan O. Talavera^a

Clinical research XVI. Differences between medians with Mann-Whitney *U* test

If you want to prove that there are differences between two groups with quantitative variables with non-normal distribution, the Mann-Whitney *U* test is used. This test is based on rank differences and it is the opposing of the Student *t* test that use quantitative variables with a normal distribution. If you want to compare three or more non related groups, the Kruskal-Wallis test is applied. When two related samples are compared Wilcoxon test is the best option (a before and after maneuver comparison), when three related samples are compared, Friedman test is used. These test correspond to the parametric opposing paired *t* test and Anova, respectively.

Key words

biomedical research
statistical analysis
reproducibility of results

En la parte XV de esta serie se mostró que la *t* de Student manifiesta las diferencias de medias de dos grupos cuando la variable cuantitativa tiene distribución normal. En este artículo se aborda cómo evaluar una variable ordinal o cuantitativa de libre distribución o, en otras palabras, que no tiene distribución normal. A este grupo de pruebas se le llama “estadística no paramétrica” y la más característica de ellas es la *U* de Mann-Whitney, que contrasta las medianas de dos grupos independientes con libre distribución.

● Ejemplo

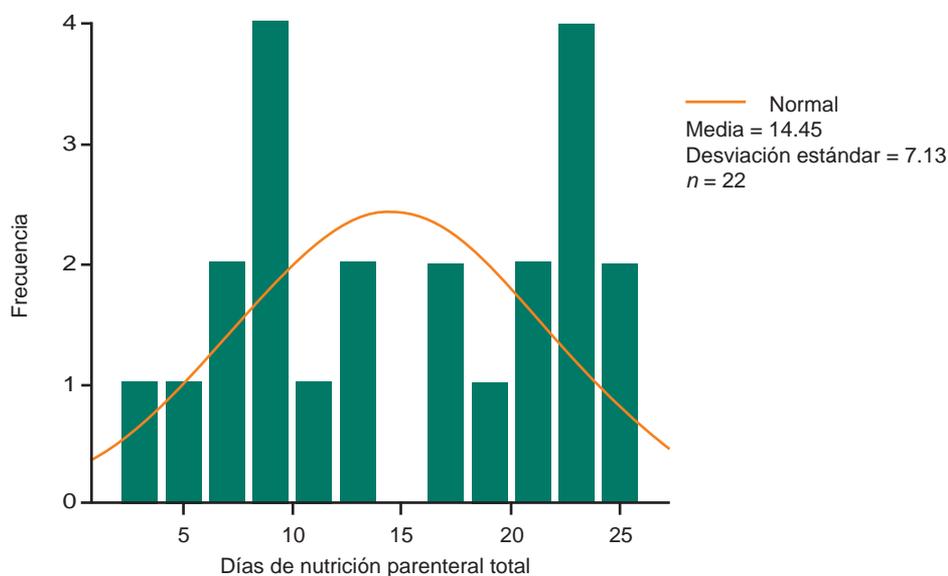
En una unidad de cuidados intensivos neonatales se realiza un estudio para saber si el tiempo de exposición a la nutrición parenteral total se relaciona con el desarrollo de colestasis. Para ello se mide el tiempo de exposición a la nutrición parenteral total en días (variable cuantitativa con libre distribución), la cual constituye la maniobra. Posteriormente se evalúa el desarrollo de la colestasis, que corresponde al desenlace (variable nominal [presente o ausente]). De tal forma, hay una variable cuantitativa de libre distribución y una variable nominal que da lugar a dos grupos.

●●● La hipótesis

El tiempo de uso de la nutrición parenteral total es mayor entre los niños que desarrollan colestasis comparado con los que no la desarrollan.

Para refutar o confirmar esta hipótesis, primero se debe conocer la distribución de la muestra. Para saber si la prueba tiene o no libre distribución se pueden usar las curvas de normalidad (figura 1) y las gráficas Q-Q (figura 2).

Figura 1 Curva de Gauss o curva de normalidad. Una curva de distribución normal debe ser simétrica y descendente a partir de la media (punto mayor), lo cual no ocurre en este ejemplo



Para demostrar que existen diferencias entre grupos independientes con variables cuantitativas que tienen libre distribución, se utiliza la U de Mann-Whitney. Esta prueba tiene su base en la diferencia de rango y es la contraparte de la t de Student que se emplea en las variables cuantitativas con distribución normal. Si se quiere comparar tres o más grupos no relacionados, la prueba que debe aplicarse es la de Kruskal-Wallis. Para contrastar dos muestras relacionadas, la prueba adecuada es la prueba de Wilcoxon (una

comparación de dos momentos: antes y después de una maniobra) y cuando se comparan tres muestras relacionadas debe recurrirse a la de Friedman. Estas pruebas corresponden a sus contrapartes paramétricas t pareada y Anova, respectivamente.

Resumen

Palabras clave

investigación biomédica
análisis estadístico
reproducibilidad de resultados

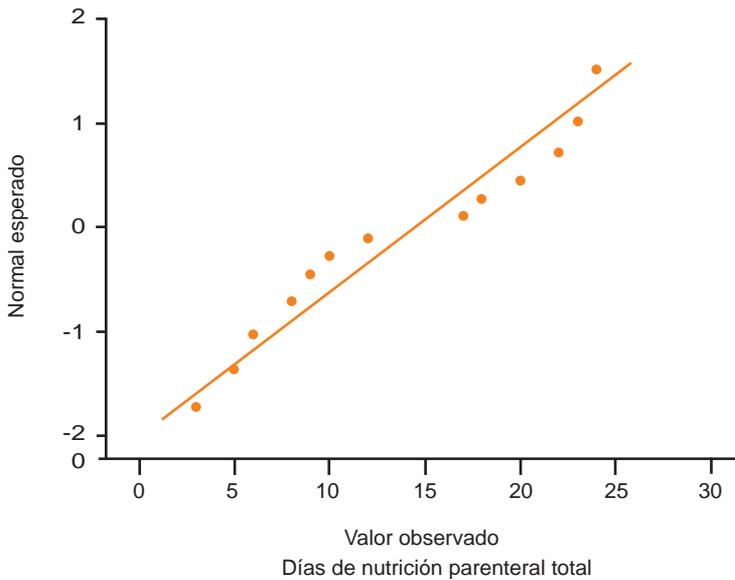


Figura 2 Gráfica Q-Q de normalidad. La línea continua representa la normalidad teórica y los puntos son los valores reales agrupados alrededor de ella. Entre más cercanos estén los puntos de la línea, mayor la posibilidad de que la muestra tenga distribución normal

Tanto la curva de normalidad como la gráfica Q-Q requieren interpretación, la cual puede ser subjetiva. Por ese motivo se calculan otros estadísticos como el sesgo y la curtosis. Para determinar normalidad, ambos valores deben estar en ± 0.5 . Si cualquiera de los dos valores, ya sea el sesgo o la curtosis, están fuera del rango de ± 0.5 , se asume que la muestra está sesgada y, por lo tanto, tiene una distribución distinta a la normal, es decir, es de libre distribución (cuadro I).

Otra manera de conocer la distribución de la muestra son las pruebas de hipótesis de Kolmogórov-Smirnov y Shapiro-Wilk. Se utiliza la primera si la muestra es mayor de 30 pacientes y la segunda, si está integrada por 30 o menos pacientes. Se asume que la muestra proviene de una población con distribución normal, por lo que la prueba de hipótesis establece lo siguiente:

$$H_0 = \text{la muestra} = \text{distribución normal} \rightarrow p > 0.05$$

$$H_a = \text{la muestra} \neq \text{distribución normal} \rightarrow p \leq 0.05$$

Cuadro I Descripción de la muestra, donde aparecen los distintos resultados de los estadísticos y en la parte de abajo el valor del sesgo y la curtosis

	Estadístico	Error estándar
Media	14.455	1.520
Intervalo de confianza para la media de 95 %		
Límite inferior	11.294	
Límite superior	17.616	
Media recortada de 5 %	14.551	
Mediana	14.5	
Varianza	50.831	
Desviación estándar	7.130	
Mínimo	3	
Máximo	24	
Rango	21	
Amplitud intercuartil	14	-0.039
Sesgo	-0.039	0.491
Curtosis	-1.592	0.953

Días de nutrición parenteral total

Cuadro II Resultado de la prueba de normalidad para días de uso de la nutrición parenteral en la UCIN

	Kolmogórov-Smirnov			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Días NPT	0.145	22	0.200*	0.904	22	0.036

El grado de libertad (gl) representa $n = 1$, por lo que la muestra se compone de 23 niños

UCIN = unidad de cuidados intensivos neonatales, Sig. = significación verdadera, NPT = nutrición parenteral total

*Límite inferior

Es decir, la muestra es de libre distribución si la p es estadísticamente significativa, como se muestra en el cuadro II (Shapiro-Wilk, $p = 0.036$). Se usa la prueba de Kolmogórov-Smirnov si la población es mayor de 30 o Shapiro-Wilk cuando la muestra es igual o menor de 30.

Una vez que conocemos que la muestra es de libre distribución, se procede a realizar la prueba de contraste de hipótesis, en este caso, la U de Mann-Whitney:

$$U_1 = R_1 - \frac{n_1(n_1 + 1)}{2}$$

$$U_2 = R_2 - \frac{n_2(n_2 + 1)}{2}$$

Donde:

U_1 = muestra con el menor número de pacientes

U_2 = tamaño de muestra mayor

R_1 y R_2 = suma de rangos para cada grupo

La designación n_1 y n_2 es arbitraria si los grupos contienen el mismo número de pacientes. Cada una de las fórmulas da un valor de U distinto (cuadro III).

El primer paso es ordenar de mayor a menor el total de la muestra (ambos grupos); a esto se le llama *rank* o clasificar.

Posteriormente, se calcula la medida de resumen (mediana). Después de ordenar de menor a mayor (5, 7, 9, 12, 18 días), los cinco pacientes del grupo sin colestasis, el que se encuentra a la mitad corresponde a la mediana (valores “rankeados” 1, 3, 4, 6 y 8), en este caso es el “rankeado” con el 4, que corresponde al sujeto con nueve días de exposición a la nutrición parenteral total. De los seis pacientes (2, 5, 7, 9, 10, 11) del grupo con colestasis, la mediana se encuentra entre el 3 y el 4, que corresponden a los valores “rankeados” con 7 y 9 y que corresponden a 15 y 24 días de nutrición parenteral total, los cuales se suman y se dividen entre dos (29/2), con lo que se obtiene una mediana de 19.5 días.

Posteriormente se calculan los valores de U_1 y U_2 :

$$U_1 = 22 - \frac{5(5 + 1)}{2} = 15$$

$$U_2 = 44 - \frac{6(6 + 1)}{2} = 21$$

Finalmente, el menor valor de U (en este caso es 15) se coteja en una tabla de valores críticos de alfa (α) de 0.05 para la prueba U de Mann-Whitney, para conocer el valor de p . En este ejemplo, p es de 0.242, es decir, aunque parece haber diferencias en esta pequeña muestra, no alcanzan a ser estadísticamente significativas.

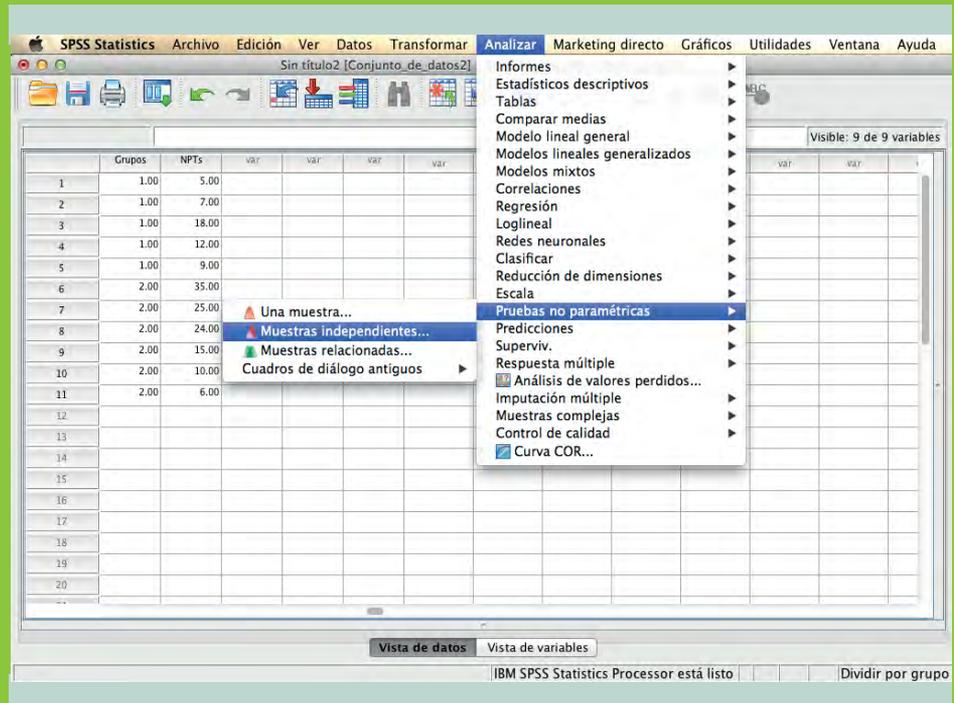
Cuadro III Orden jerárquico de los valores arrojados por los 11 pacientes estudiados en la Unidad de Cuidados Intensivos Neonatales

Sin colestasis $n_1 = 5$	Rank (orden)	Con colestasis $n_2 = 6$	Rank (orden)
5	1	35	11
7	3	25	10
18	8	24	9
12	6	15	7
9	4	10	5
		6	2
Mediana = 9		Mediana = 19.5	
R = 22		R = 44	

Pasos en el programa SPSS

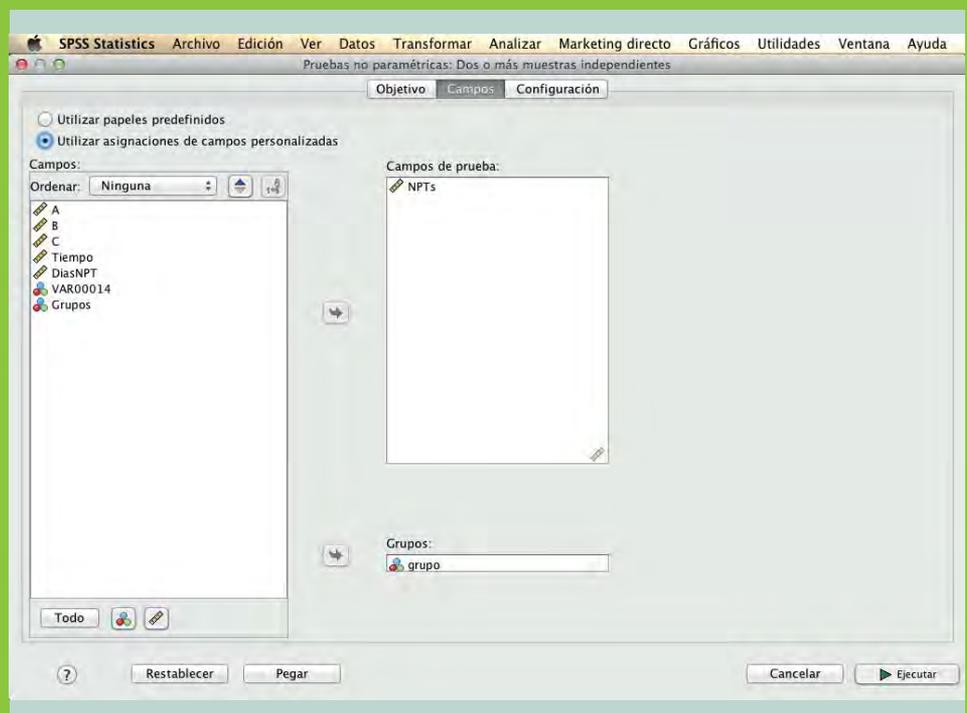
1

En el menú se selecciona la opción “Analizar” y en esta, “Pruebas no paramétricas”. Una vez ahí, se elige “Muestras independientes”.



2

En la siguiente pantalla se solicita ingresar los grupos por contrastar, que en este caso son dos: con colestasis y sin colestasis, y una variable cualitativa, que en este caso es el tiempo de uso de la nutrición parenteral total.



3

Finalmente, se presiona “Ejecutar” y se obtiene el resultado.

	Hipótesis nula	Test	Sig.	Decisión
1	Las medianas de NPTs son las mismas entre las categorías de grupo.	Prueba de medianas de muestras independientes	.242 ^{1,2}	Retener la hipótesis nula.
2	La distribución de NPTs es la misma entre las categorías de grupo.	Prueba U de Mann-Whitney de muestras independientes	.177 ¹	Retener la hipótesis nula.

Se muestran las significancias asintóticas. El nivel de significancia es .05.

¹ Se muestra la significancia exacta para esta prueba.

² Sig. exacta de Fisher

Comentarios

La prueba *U* de Mann-Whitney permite comparar dos medianas, ya sea que provengan de una variable ordinal o de una cuantitativa con libre distribución. Para llegar a esta conclusión será necesario explorar algunos métodos gráficos (imagen de una curva de supervivencia o gráfica Q-Q de normalidad) o algebraicos (sesgo y curtosis o Kolmogórov-Smirnov o Shapiro-Wilk).

Es necesario hacer énfasis en que la *U* de Mann-Whitney es específica para comparar dos muestras

independientes, pero si se trata de tres muestras deberá utilizarse la prueba de Kruskal-Wallis y si se trata de dos muestras relacionadas, la de Wilcoxon; para más de dos muestras relacionadas, la prueba idónea es la de Friedman.

Declaración de conflicto de interés: los autores han completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no fue reportado alguno en relación con este artículo.

^aCentro de Adiestramiento en Investigación Clínica, Coordinación de Investigación en Salud, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Distrito Federal, México

Comunicación con: Rodolfo Rivas-Ruiz
Correo electrónico: rivastrodolfo@gmail.com

Lecturas recomendadas

1. Armitage P, Berry G, Matthews JNS. Statistical methods in medical research. Fourth edition. Massachusetts: Blackwell; 2002.
2. Bland M. Introduction to medical statistics. Third edition. New York: Oxford University Press; 2000.
3. Dawson B, Trapp RG. Basic and clinical biostatistics. Fourth edition. New York: Lange Medical Books-McGraw-Hill; 2004.
4. Feinstein AR. Clinical biostatistics. St. Louis: Mosby; 1977.
5. Feinstein AR. Clinical epidemiology: the architecture of clinical research. Philadelphia: Saunders; 1985.
6. Feinstein AR. Multivariable analysis: an introduction. New Haven: Yale University Press; 1996.
7. Le Chap T. Introductory biostatistics. New Jersey: Wiley-Interscience; 2003. Texto libre en [http://www.hstathome.com/tjziyuan/Introductory%20Biostatistics%20Le%20C.T.%20\(Wiley,%202003\)\(T\)\(551s\).pdf](http://www.hstathome.com/tjziyuan/Introductory%20Biostatistics%20Le%20C.T.%20(Wiley,%202003)(T)(551s).pdf)
8. Peat J, Barton B. Medical statistics. A guide to data analysis and critical appraisal. Massachusetts: Blackwell; 2005.
9. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Third edition. Upper Saddle River, NJ: Pearson, Prentice Hall; 2009.
10. Portney LG, Watkins MP. Foundations of clinical research: applications to practice. Norwalk, Conn.: Appleton and Lange; 1993.
11. Rivas-Ruiz R, Pérez-Rodríguez M, Talavera JO. Del juicio clínico al modelo estadístico. Diferencia de medias. Prueba *t* de Student. Rev Med

- Inst Mex Seguro Soc. 2013;51(3):300-3. Texto libre en http://revistamedica.imss.gob.mx/index.php?option=com_multicategories&view=article&id=2077:51-3-11-investigacion&catid=785:temas-de-actualidad&Itemid=775
12. Talavera JO, Rivas-Ruiz R. Investigación clínica IV. Pertinencia de la prueba estadística. Rev Med Inst Mex Seguro Soc. 2011;49(4):401-5. Texto libre en http://revistamedica.imss.gob.mx/index.php?option=com_multicategories&view=article&id=1432:investigacion-clinica-iv-pertinencia-de-la-prueba-estadistica&catid=508:reportes-breves&Itemid=672
13. Talavera JO, Rivas-Ruiz R. Investigación clínica V. Tamaño de muestra. Rev Med Inst Mex Seguro Soc. 2011;49(5):517-22. Texto libre en http://revistamedica.imss.gob.mx/index.php?option=com_multicategories&view=article&id=1452:investigacion-clinica-v-tamano-de-muestra&catid=514:temas-de-actualidad&Itemid=676
14. Talavera JO, Wachter-Rodarte NH, Rivas-Ruiz R. Investigación clínica III. Estudios de causalidad. Rev Med Inst Mex Seguro Soc. 2011;49(3): 289-94. Texto libre en http://revistamedica.imss.gob.mx/index.php?option=com_multicategories&view=article&id=1410:investigacion-clinica-iii-edtudios-de-causalidad&catid=501:temas-de-actualidad&Itemid=664



Investigación clínica XVIII

Del juicio clínico al modelo de regresión lineal

Lino Palacios-Cruz,^{a,b} Marcela Pérez,^b Rodolfo Rivas-Ruiz,^b Juan O. Talavera^b

Cada peculiaridad en un hombre es compartida por sus descendientes, pero en promedio, en un grado menor

SIR FRANCIS GALTON, 1886

From clinical judgment to linear regression model

When we think about mathematical models, such as linear regression model, we think that these terms are only used by those engaged in research, a notion that is far from the truth. Legendre described the first mathematical model in 1805, and Galton introduced the formal term in 1886. Linear regression is one of the most commonly used regression models in clinical practice. It is useful to predict or show the relationship between two or more variables as long as the dependent variable is quantitative and has normal distribution. Stated in another way, the regression is used to predict a measure based on the knowledge of at least one other variable. Linear regression has as its first objective to determine the slope or inclination of the regression line: $Y = a + bx$, where "a" is the intercept or regression constant and it is equivalent to "Y" value when "X" equals 0 and "b" (also called slope) indicates the increase or decrease that occurs when the variable "x" increases or decreases in one unit. In the regression line, "b" is called regression coefficient. The coefficient of determination (R^2) indicates the importance of independent variables in the outcome.

Key words

linear models
models, statistical
statistics

El término *modelo matemático* nos remite a conceptos que solo atañen a quienes se dedican a investigar, noción que dista mucho de la realidad. En la práctica clínica y en la vida diaria realizamos asociaciones o predicciones que nos ayudan en nuestro desempeño cotidiano, por ejemplo, cuando evaluamos a un adolescente con un trastorno por uso de alcohol y deseamos inferir el efecto de los factores biológicos y medioambientales (y también indirectamente el pronóstico a mediano y largo plazo) mediante la asociación de la edad al inicio del consumo de alcohol y el número de familiares con el mismo hábito. En casa, no es infrecuente tratar de predecir la cantidad de un producto que debemos comprar para una semana si hay menos integrantes de la familia que los habituales.

Este método es valioso en escenarios clínicos o económicos, como cuando deseamos evaluar variables que, por su costo o dificultad para su obtención, requieren alguna aproximación clínica previa, como la estimación de la densidad mineral ósea a partir de la medición de peso, talla y la ultrasonografía ósea.

En este punto, por más que el lector juró "no volver a tratar con las matemáticas y alejarse lo más posible de ellas", tal vez comienza a sospechar que la relación es útil, ya que de manera natural aplica modelos matemáticos simples y complejos, en un nivel de menor conciencia pero con un nivel equiparable de utilidad.

Los métodos estadísticos, como la regresión lineal, permiten predecir o disminuir esa incertidumbre.¹ El análisis de regresión se define como "el estudio de la dependencia" o de cómo una respuesta o variable depende de uno o más predictores o variables independientes. Al considerar este modelo en un proyecto de investigación o análisis de la información es importante tomar en cuenta dos aspectos básicos:

- Que la dependencia de la respuesta sobre los predictores se lleva a cabo mediante el promedio, por lo tanto, se requiere que esta variable tenga una distribución normal.
- Que el promedio de la variable dependiente dadas las variables independientes es una función lineal, es decir, la variable dependiente se incrementa o disminuye conforme se incrementan o disminuyen los valores de las variables independientes o predictoras.² Dicho de otra manera: debe existir una relación en la que el incremento o disminución de una variable sea proporcional en cada punto.

Se considera *regresión lineal simple* si se relacionan solo dos variables, de las cuales la dependiente es cuantitativa. Cuando se utilizan dos o más variables para predecir una variable cuantitativa se considera *regresión lineal múltiple*. Las variables independientes pueden combinar variables cuantitativas y cualitativas.

Pensamos que los modelos matemáticos, como la regresión lineal, son conceptos que solo atañen a quienes se dedican a investigar, noción que dista de la realidad. La primera descripción de un modelo matemático fue realizada por Legendré, en 1805, y la introducción formal del término fue hecha por Galton, en 1886. La regresión lineal es útil para predecir la relación entre dos o más variables, siempre y cuando la variable dependiente sea cuantitativa y cuente con una distribución normal. Su desarrollo tiene como primer objetivo determinar la pendiente o inclinación de la línea de regresión: $Y = a + bx$, donde "a" es la "constante de regresión" que equivale al valor de "Y"

cuando "x" es igual a 0 y "b", también llamada pendiente de la recta, indica el incremento o decremento que se produce en "Y" cuando la variable "x" aumenta o disminuye una unidad. En la línea de regresión, "b" recibe también el nombre de coeficiente de regresión. El coeficiente de determinación (R^2) define la magnitud de la capacidad para predecir el efecto de las variables independientes sobre el resultado.

Resumen

Palabras clave

modelos lineales
modelos estadísticos
estadística

Probablemente, la regresión lineal es, junto con la regresión logística, el modelo de regresión más aplicado, tanto en las investigaciones de las ciencias naturales y sociales como en la vida diaria.

Historia de la regresión lineal

Si bien la primera descripción documentada sobre un método de regresión lineal fue publicada por Legendré en 1805, en el método de mínimos cuadrados con el que abordaba una versión del teorema de Gauss-Márkov,²⁻⁴ fue sir Francis Galton, médico y primo de Charles Darwin, quien introdujo el término *regresión*, en su artículo "Regression towards mediocrity in hereditary stature", publicado en 1886 en el *Journal of the Anthropological Institute*⁵ y que menciona de nuevo en su libro *Natural Inheritance*, de 1889.⁶ En ese trabajo clásico, Galton centró su descripción en los rasgos físicos de los descendientes (variable dependiente) a partir de los rasgos de sus padres (variable independiente). Analizó la altura de 205 padres y 930 hijos adultos a partir de sus registros familiares y llegó a la conclusión de que los padres muy altos tenían una tendencia a tener hijos que heredaban parte de esta altura, pero que se revelaba también una tendencia a regresar a la estatura media. A partir de estas observaciones, Galton señaló esta tendencia bajo la "ley de la regresión universal". Al final del mismo siglo XIX, Pearson y Yule aportaron muchas de las nociones modernas acerca de la correlación, que han contribuido al metalenguaje que nos permite el entendimiento del fenómeno de la dependencia entre variables.

Teoría y conceptos de la regresión lineal

Como se ha señalado, la regresión se utiliza para predecir una medida o variable dependiente (también llamada de desenlace o variable "y") basándonos en el

conocimiento de al menos otra variable independiente (o variable relacionada con la maniobra o variable "x")⁷ y un término aleatorio ε .⁸

El proceso de regresión lineal tiene como primer paso determinar la pendiente o inclinación de la línea de regresión, cuya representación algebraica para la regresión lineal simple es de la siguiente forma:¹

$$E(Y/X) = a + bx$$

Donde:

El estimador de "Y" dado un valor de "X" es igual a $a + b$ que multiplica "x", asumiendo que la distribución de "Y" para una "x" determinada es normal y, además, que las varianzas de ambas variables son homogéneas, fenómeno conocido como *homocedasticidad*.^{1,9,10}

La manera más popular de la representación matemática de la regresión lineal simple es como sigue:

$$y_i = B_0 + B_1 X_i + \varepsilon$$

Donde:

y_i = variable explicada o dependiente. La "y", a diferencia de la "Y", que es un valor real dentro de la población, es el indicador de "Y" o el valor estimado a partir de una muestra que trata de predecir "Y".

B_0 = intercepto, ordenada al origen o constante de regresión, que es la altura a la que la recta corta al eje "Y", equivalente al valor de "y" cuando "x" es igual a cero.

B_1 = parámetro que mide la influencia que tienen las variables explicativas sobre la variable explicada o dependiente.

● Ejemplo

En un grupo de 40 adultos jóvenes (67.5 % mujeres y edad promedio de 31.5 ± 4.71) que acuden a un servicio de salud mental, se desea predecir la edad de inicio del trastorno depresivo mayor (11 ± 3.63 años), a partir de la edad de inicio del trastorno por déficit de atención con hiperactividad (10.3 ± 1.63 años) (cuadro I). Al trazar para cada individuo el valor que obtienen en la edad de inicio para depresión (en el eje de las "Y") y para el trastorno por déficit de atención con hiperactividad (en el eje de las "X"), obtendríamos una gráfica similar a la de la figura 1.

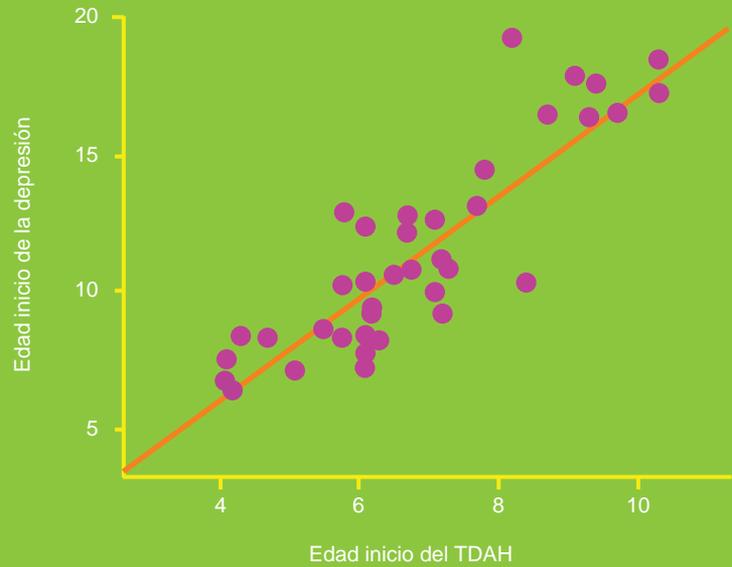


Figura 1 Regresión lineal para predecir la edad de inicio de depresión a partir de la edad de inicio del trastorno por déficit de atención con hiperactividad

Cuadro I Edad de inicio (en años) de los trastornos depresivo mayor y por déficit de atención con hiperactividad en 40 adultos jóvenes

TDM				TDAH			
Paciente	Edad	Paciente	Edad	Paciente	Edad	Paciente	Edad
1	19.2	21	10.3	1	8.2	21	6.1
2	6.7	22	9.9	2	4.1	22	7.1
3	8.3	23	9.2	3	4.7	23	7.2
4	6.4	24	16.3	4	4.2	24	9.3
5	7.1	25	10.2	5	5.1	25	5.8
6	11.2	26	16.4	6	7.2	26	8.7
7	14.4	27	17.5	7	7.8	27	9.4
8	12.7	28	9.3	8	6.7	28	6.2
9	9.1	29	8.2	9	8.1	29	6.3
10	13.1	30	11.1	10	7.7	30	7.2
11	10.6	31	12.6	11	6.5	31	7.1
12	8.3	32	10.8	12	4.3	32	6.8
13	7.5	33	9.1	13	4.1	33	6.2
14	17.8	34	7.7	14	9.1	34	6.1
15	10.3	35	8.3	15	8.4	35	5.8
16	17.2	36	8.6	16	10.3	36	5.5
17	16.4	37	8.4	17	9.7	37	6.1
18	12.3	38	10.8	18	6.1	38	7.3
19	12.8	39	7.3	19	5.8	39	6.1
20	12.1	40	18.4	20	6.7	40	10.3

TDM = trastorno depresivo mayor, TDAH = trastorno por déficit de atención con hiperactividad

Para la regresión lineal simple, B_1 corresponde a la pendiente de la línea y da la proporción de cambio en “y” por cada unidad de cambio en “x”. Al término, a B_1 se le conoce como *coeficiente de regresión*. La positividad (relación directa) o negatividad (relación inversa) de la línea de regresión depende de B_1 : si el valor de B_1 es positivo, “y” aumenta conforme “x” aumenta; si el valor de B_1 es negativo, “y” disminuye conforme “x” aumenta. Finalmente, si b es igual a cero, no hay cambios en “y”, por lo tanto, la pendiente se mantiene en forma horizontal sin predecir “y” a partir de “x”.¹

La regresión lineal simple se puede ampliar para considerar simultáneamente los efectos de dos o más variables explicativas (es decir, regresión lineal múltiple). Existen otros modelos de regresión, como cuando

la variable dependiente tiene un desenlace binario (regresión logística) o se trata de un desenlace binario en el transcurso del tiempo (riesgos proporcionales de Cox) o de recuento (regresión de Poisson).⁹ En una regresión lineal múltiple podría expresarse como:

$$Y_i = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \dots + B_pX_p + \varepsilon$$

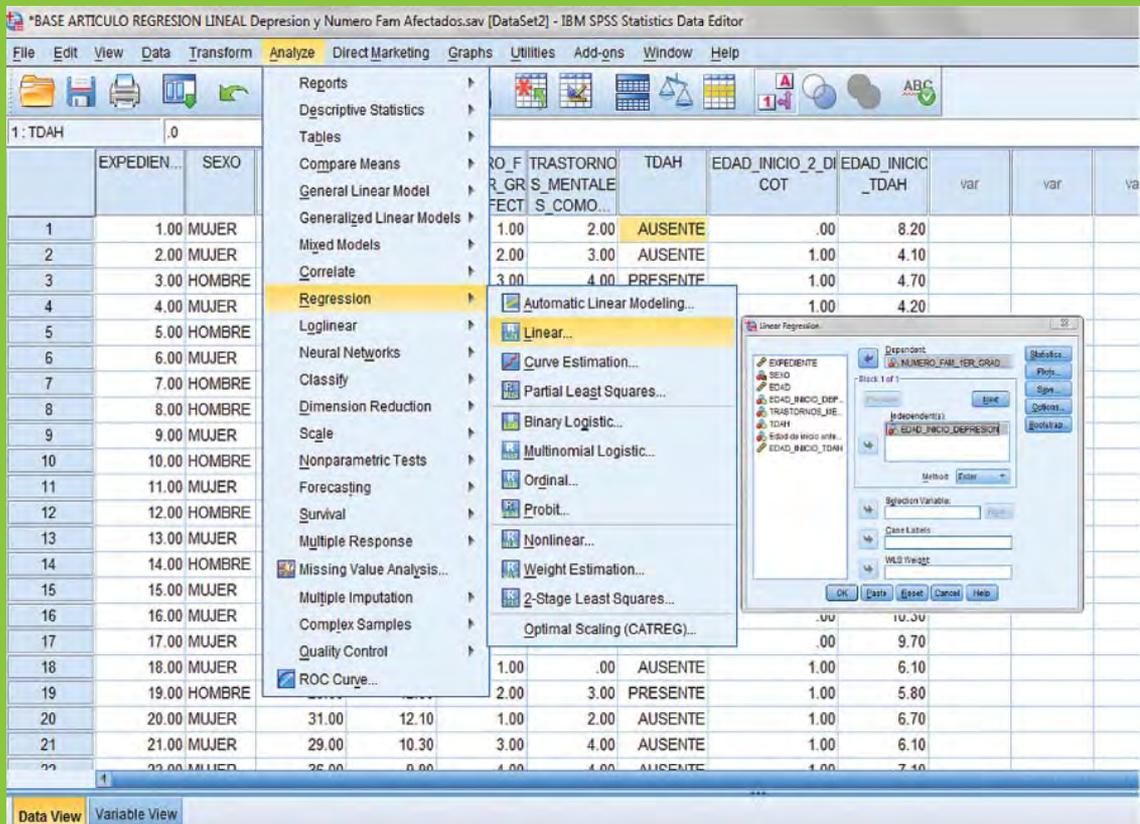
De manera similar, Y_i es la variable explicada o dependiente.

B_1, B_2, B_3, \dots son los parámetros que miden la influencia que tienen las variables explicativas sobre la variable explicada o dependiente. Las B_i ($i > 0$) son los parámetros respectivos de cada variable independiente y p es el número de parámetros independientes por considerar en la regresión.

●●● De regreso al ejemplo

Ahora bien, si se desea determinar si la edad de inicio del trastorno depresivo mayor (el grupo podría dividirse en dos subgrupos: sujetos que iniciaron con el trastorno antes de los 12 años y sujetos cuyo inicio del trastorno fue a partir de los 13 años, o bien, dejar la variable a nivel dimensional) predice el número de familiares en primer grado afectados con el mismo trastorno, en la figura 2 se observan los pasos para realizar el análisis en un programa estadístico.

Figura 2 Pasos para realizar una regresión lineal en el programa SPSS



En la predicción mediante la regresión lineal, un paso importante es la suma de los mínimos cuadrados, que corresponden a las cantidades que minimizan la suma de cuadrados de la varianza $(y - Y)^2$, ecuación que representa la recta con la menor distancia de “y” a “Y” (distancia entre un valor estimado y un valor real), pero elevada al cuadrado con el fin de no obtener un valor de 0, dado que “Y” se distribuye por igual forma a cada lado de la línea de regresión.^{1,11}

En el estudio de la distancia del valor real y el valor estimado es necesario calcular un valor mínimo y un valor máximo del coeficiente de regresión, es decir, determinar el intervalo de confianza de 95 %, que proporciona las desviaciones explicadas por la pendiente de regresión.^{1,7,12}

Al realizar la regresión lineal con los datos del ejemplo y al procesarlos con el programa ahora llamado IBM SPSS® (figura 3), podemos observar que

la suma de cuadrados del modelo de regresión — es decir, el nivel de fluctuación de la variable Y_t que el modelo es capaz de explicar— tiene un valor de 21.24 y la suma de cuadrados de los modelos residuales —es decir, el indicador del nivel de error del modelo— es de 23.16 (o el porcentaje no explicado por el modelo). El valor de la constante, es decir el valor de “Y” cuando “X” es igual a 0, es de 4.12 y el coeficiente de regresión es de -0.204 , con un intervalo de confianza de 95 % que va de -0.274 a -0.134 . Es decir, mediante este modelo de regresión puede establecerse que a menor edad al iniciar el trastorno depresivo mayor, existe una mayor probabilidad de tener un mayor número de familiares en primer grado con el mismo trastorno.

Si bien mediante la línea de regresión se intenta predecir el valor de una variable a partir de otra, esta no proporciona en forma directa el porcentaje de la asociación de “y” (en el ejemplo, número de familia-

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.692 ^a	.478	.465	.78070

a. Predictors: (Constant), EDAD_INICIO_DEPRESION

- Suma de cuadrados de la regresión (**sum of squares, regression**): indica qué tanta variación de la variable dependiente explica nuestro modelo.
- Suma de cuadrados de los residuales (**sum of squares, residual**): indica qué tanta variación de la variable dependiente no explica nuestro modelo.

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	21.239	1	21.239	34.848	.000 ^b
	Residual	23.161	38	.609		
	Total	44.400	39			

a. Dependent Variable: NUMERO_FAM_1ER_GRADO_AFECT

b. Predictors: (Constant), EDAD_INICIO_DEPRESION

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.119	.412		10.003	.000
	EDAD_INICIO_DEPRESION	-.204	.035	-.692	-5.903	.000

a. Dependent Variable: NUMERO_FAM_1ER_GRADO_AFECT

Figura 3 Resultados de la regresión lineal

res en primer grado con trastorno depresivo mayor) a partir de “x” (en el ejemplo, edad de inicio del trastorno depresivo mayor). La intensidad de la asociación o el porcentaje de explicación del modelo se define con el coeficiente de determinación (R^2), que puede ser corregido o no y que equivale al cuadrado del coeficiente de correlación “R” (figura 3), que en nuestro modelo fue igual a 0.692.

La R^2 corregida para este modelo fue de 0.465, cuya traducción es que el número de familiares en primer grado que podrían tener el trastorno depresivo mayor se relaciona en 46.5 % con la edad de inicio de ese trastorno en el caso índice; en 53.5 % de los casos se debe a otros factores no incluidos en este modelo ($1 - R^2$).

Uno de los últimos pasos de este método de análisis es el establecimiento de la significación de la curva de regresión mediante la prueba de hipótesis, en la que se supone que el coeficiente de regresión no es igual a 0. Si bien el programa SPSS permite obtener automáticamente este resultado, anteriormente se debía analizar la varianza de la regresión, en la que si el coeficiente de regresión no es igual a 0, el valor de “F” observado es mayor que el valor crítico de “F”, lo que se corresponde con un valor de $p < 0.05$. De esta forma, se rechaza la hipótesis nula, se acepta la hipótesis alterna y, de esa manera, se determina que la pendiente sí permite predecir “y” a partir de “x”, es decir, que existe significación estadística.

Es importante recordar que la predicción de la variable dependiente a partir de una o más variables independientes no significa causalidad y que esta solo deberá considerarse si se cumplen las condiciones enunciadas en mayo de 1965 por Austin Bradford Hill: plausibilidad, especificidad, temporabilidad, etcétera.¹³

Conclusiones

El análisis de regresión es solo una herramienta más dentro de las opciones para que el clínico y el investigador se acerquen a la naturaleza de los resultados.^{14,15} Los aspectos importantes que deben recordarse al momento de realizar y leer correctamente los resultados del análisis de regresión lineal son:

- La variable dependiente debe ser continua.
- La o las variables independientes pueden ser continuas o categóricas.
- El intercepto.
- El coeficiente de regresión.
- El coeficiente de determinación (R^2), importante para definir la magnitud de la relación de la o las variables predictoras sobre la variable resultante o predicha.
- El intervalo de confianza.
- El valor de “F” del análisis de la varianza de la regresión.

^aInstituto Nacional de Psiquiatría “Ramón de la Fuente Muñiz”, Secretaría de Salud, Distrito Federal, México

^bCentro de Adiestramiento en Investigación Clínica, Coordinación de Investigación en Salud, Centro Médico

Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Distrito Federal, México

Correspondencia: Lino Palacios-Cruz
Correo electrónico: palacioslino@gmail.com

Referencias

1. Talavera-Piña JO, Antonio-Ocampo A, Castellanos-Olivares A, Wachter-Rodarte NH. Regresión lineal simple. *Rev Med IMSS*. 1995;33(3):347-51.
2. Weisberg S. Linear hypothesis: regression (basics). En: Neil JS, Paul BB, editores. *International encyclopedia of the social & behavioral sciences*. Oxford: Pergamon; 2001. p. 8884-8.
3. Wikipedia. La Enclopedia Libre. Regresión lineal. [Consultado el 2 de septiembre de 2013]. Disponible en <http://es.wikipedia.org/w/index.php?oldid=67582255>.
4. Lam GK. On the unit of dose equivalent and the linear hypothesis. *Health Phys*. 1989;57(3):495-6.
5. Galton F. Regression towards mediocrity in hereditary stature. *J Anthropol Inst*. 1886;15:246-63.
6. Galton F. *Natural inheritance*. London-New York: MacMillan; 1889.
7. Pearce J. Regression, linear and nonlinear. En: Rob K, Nigel T, editores. *International encyclopedia of human geography*. Oxford: Elsevier; 2009. p. 302-8.
8. Canavos GC. *Probabilidad y estadística. Aplicaciones y métodos*. México: McGraw-Hill; 1988.
9. Devore JL. *Probabilidad y estadística para ingeniería y ciencias*. México: International Thomson Editores; 2005.
10. Dawson P, Trapp R. *Bioestadística médica*. México: Manual Moderno; 1990. p. 239-365.
11. Seber GAF. *The linear hypothesis: a general theory*. London: Griffin; 1966.
12. Raftery AE, Madigan D, Hoeting JA. Bayesian model averaging for linear regression models. *J Am Stat Assoc*. 1997;92(437):179-91.
13. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58(5):295-300.
14. Talavera JO, Rivas-Ruiz R. Investigación clínica IV. Pertinencia de la prueba estadística. *Rev Med Inst Mex Seguro Soc*. 2011;49(4):401-5.
15. Talavera JO, Rivas-Ruiz R. Investigación clínica XIV. Del juicio clínico al modelo estadístico. *Rev Med Inst Mex Seguro Soc*. 2013;51(5):170-5.

Investigación clínica XIX

Del juicio clínico al análisis de covarianza

Marcela Pérez-Rodríguez,^a Lino Palacios-Cruz,^b Jorge Moreno,^a Rodolfo Rivas-Ruiz,^a Juan O. Talavera^a

La selección natural es un mecanismo para generar un altísimo grado de improbabilidad

SIR RONALD A. FISHER, 1947

Clinical research XIX. From clinical judgment to analysis of covariance

The analysis of covariance (ANCOVA) is based on the general linear models. This technique involves a regression model, often multiple, in which the outcome is presented as a continuous variable, the independent variables are qualitative or are introduced into the model as dummy or dichotomous variables, and factors for which adjustment is required (covariates) can be in any measurement level (i.e. nominal, ordinal or continuous). The maneuvers can be entered into the model as 1) fixed effects, or 2) random effects. The difference between fixed effects and random effects depends on the type of information we want from the analysis of the effects. ANCOVA effect separates the independent variables from the effect of co-variables, i.e., corrects the dependent variable eliminating the influence of covariates, given that these variables change in conjunction with maneuvers or treatments, affecting the outcome variable. ANCOVA should be done only if it meets three assumptions: 1) the relationship between the covariate and the outcome is linear, 2) there is homogeneity of slopes, and 3) the covariate and the independent variable are independent from each other.

Key words

analysis of covariance
factor analysis, statistical

Palabras clave

Análisis de covarianza
Análisis factorial

En la vida diaria es común tener un pensamiento lineal, es decir, que asumamos que un factor influye directamente sobre un resultado; sin embargo, esto no necesariamente sucede así. De esta forma si queremos saber el verdadero efecto que tiene, por ejemplo, un jefe amable y considerado en la productividad de sus empleados, tenemos que hacer un ajuste a partir de la presencia de diferentes salarios

En este ejemplo podríamos utilizar el análisis de covarianza (ANCOVA) el cual es parte de los modelos generales lineales y utiliza el análisis de varianza y el de regresión para eliminar la variabilidad que provoca una covariable (condición basal o maniobra periférica) sobre el desenlace (que siempre es una variable cuantitativa), para estimar el efecto de la maniobra principal. El ANCOVA fue creado por sir Ronald Aylmer Fisher y fue publicado por primera vez en 1947 en *Biometrics*, en un trabajo titulado “The analysis of covariance method for the relation between a part and the whole”. Esta aportación se hizo formalmente famosa en 1948 al ser publicada en el libro *Statistical methods for research workers*.^{1,2}

Originalmente, el ANCOVA¹ se desarrolló para reducir el error de la varianza en experimentos aleatorizados, lo cual aumentaba tanto el poder estadístico de la prueba de hipótesis (a través de la inclusión de covariables al reducir la variabilidad) como la precisión en la estimación de los efectos.³ En el ANCOVA se consideran tres tipos de variables:³

- La o las variables independientes o maniobras (con las que se representan las condiciones experimentales o grupos que se quieren comparar).
- Las variables independientes o covariables (que representan las variables que se quiere controlar).
- La variable de desenlace (que es aquella sobre la cual actúan los otros dos tipos de variables).

Esta técnica involucra un modelo de regresión múltiple en el que el desenlace se presenta como variable continua, la maniobra (variable cualitativa) es introducida al modelo como variable *dummy* o dicotómica y los factores por los que se requiere ajustar el análisis (covariables) pueden estar en cualquier tipo de medidas (nominales, ordinales o continuas). Las maniobras pueden ser ingresadas al modelo como factores o efectos fijos o factores o efectos aleatorios. La diferencia entre manejar las maniobras como factor fijo o aleatorio depende del tipo de información que queremos del análisis de los efectos. En los efectos fijos, usualmente nos interesa hacer comparaciones explícitas entre un grupo y otro. Por ejemplo, cuando queremos comparar sujetos con distinto índice de masa corporal (IMC) para ver su asociación con la tensión arterial, comparamos el grupo de peso normal, contra el grupo que

El análisis de covarianza (ANCOVA) parte de los modelos generales lineales. Esta técnica involucra un modelo de regresión, muchas veces múltiple, en el que el desenlace se presenta como variable continua, la o las maniobras son variables cualitativas que son introducidas al modelo como variables *dummy* o dicotómicas y los factores por los que se requiere ajustar el análisis (covariables) pueden estar en cualquier nivel de medición (nominal, ordinal o continuo). Las maniobras pueden ser ingresadas al modelo como factores (o efectos) fijos o factores (o efectos) aleatorios. La diferencia entre manejar las maniobras como factor

fijo o factor aleatorio depende del tipo de información que se busca en el análisis de los efectos. El ANCOVA separa el efecto de las maniobras del efecto de las covariables, es decir, corrige la variable de respuesta eliminando la influencia de las covariables por el hecho de que estas varían conjuntamente con las maniobras o tratamientos, lo cual afecta la variable de desenlace. El ANCOVA solo debe ser realizado si se cumplen tres supuestos: 1) la relación entre la covariable y el desenlace es lineal, 2) existe homogeneidad de las pendientes y 3) se comprueba la independencia entre la covariable y la maniobra o variable independiente.

Resumen

conforman los que tienen sobrepeso, obesidad y desnutrición. En este caso tratamos la maniobra observacional (IMC) como un efecto fijo. Aun cuando nuestro interés no está específicamente en una variable, existe, sin embargo, la posibilidad de que esta agregue variación al modelo; entonces esta variable debería ser tratada como una variable de efectos aleatorios.

Para aclarar este concepto, imaginemos que queremos estudiar el efecto del género sobre las horas de participación en un programa de ejercicio en una muestra de 10 estados de la república; en este ejemplo, la variable de agrupación (que sería el estado de la república) debería ser tratada como factor aleatorio, ya que la variable dependiente o de desenlace es la cantidad de horas de participación en el programa de ejercicio y la variable independiente o maniobra es el género. Incluir esta variable de efectos aleatorios puede mejorar la varianza explicada por el género.

El ANCOVA separa el efecto de las maniobras del efecto de las covariables, es decir, corrige la variable de respuesta eliminando la influencia de las covariables por el hecho de que estas varían conjuntamente con las maniobras o tratamientos, lo cual afecta la variable de desenlace. Por ejemplo, supongamos que queremos estudiar el efecto del nivel de hierro en las embarazadas (anémicas o no anémicas) sobre el peso del recién nacido; sin embargo, la ganancia de peso durante el embarazo no se pudo controlar durante el estudio y se sabe que la ganancia de peso durante la gestación pudiera afectar también el peso del recién nacido. En este caso sería deseable separar el efecto del nivel de hierro del efecto de la ganancia de peso durante el embarazo sobre el peso del recién nacido, es decir, corregir el efecto del nivel de hierro al eliminar la influencia de la ganancia de peso.

En el análisis de la covarianza se realiza el siguiente modelo de regresión:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + E$$

Donde:

Y = desenlace (peso del recién nacido)

X = covariable (ganancia de peso en el embarazo).

Z = variable dicotómica que representa los grupos que se van a comparar (anémica y no anémica) ($Z = 0$ para un grupo y $Z = 1$ para otro grupo, etcétera).

E = error.

De acuerdo con este modelo, las medias ajustadas para los grupos se definen como los valores predichos obtenidos al evaluar el modelo con $Z = 0$ y $Z = 1$, cuando se asume artificialmente que X tiene el mismo promedio general de la covariable para todos los grupos (por ejemplo, se asume que todas las participantes ganaron la misma cantidad de peso durante el embarazo). Este procedimiento es equivalente a asumir una distribución común de las covariables que esté basada en la muestra combinada de todos los grupos. Los promedios corregidos por las covariables se llaman medias mínimo-cuadráticas corregidas (LSMEANS, por sus siglas en inglés —*least-squares means*—) y se calculan de la siguiente manera:

$$\hat{Y} = \bar{Y} - b_{yx} (X_i - \bar{X})$$

Después de realizar el modelo de regresión y calcular las medias ajustadas, se realiza una prueba de hipótesis de F ($H_0: \beta_2 = 0$) para determinar si dichas medias ajustadas son significativamente diferentes.

El ANCOVA solo debe ser realizado si se cumplen tres supuestos (cuadro I):

- La relación entre la covariable y el desenlace es lineal
- Existe homogeneidad de las pendientes.
- Se comprueba la independencia entre la covariable y la maniobra o tratamiento.

Cuadro I Supuestos y consideraciones del ANCOVA

Supuestos*
1. <i>La relación entre X y Y debe ser lineal.</i> Este supuesto se prueba si se hace un análisis de regresión lineal simple entre la covariable (X) y la respuesta (Y). La regresión debe ser significativa, es decir, con $p < 0.05$.
2. <i>Homogeneidad de las pendientes.</i> Las pendientes de todos los grupos de tratamiento deben ser iguales o aproximadamente iguales; es decir, no hay interacción entre las variables de estudio y las covariables. Este supuesto se prueba verificando que la interacción del tratamiento por la covariable no sea significativa.
3. <i>Supuesto de independencia entre la covariable (X) y el tratamiento.</i> Este supuesto se prueba realizando un ANOVA en el que la covariable (X) sea declarada variable dependiente y los tratamientos sean declarados como variables independientes. La prueba de F debe resultar no significativa.
Consideraciones
1. Teóricamente no hay límites para el número de covariables que pueden usarse, pero la práctica ha demostrado que 4 o 5 son suficientes; más allá de este número, se generan problemas de colinearidad.
2. Las covariables son variables independientes que afectan la respuesta, pero sobre las cuales no se ha podido ejercer control por criterios de selección o por alguna otra razón.
3. Las covariables deben ser medidas en escala de razón, intervalar o nominal.
4. Las covariables no deben estar relacionadas con las maniobras o condiciones experimentales.
5. Las covariables deben estar relacionadas con el desenlace (variable dependiente o de respuesta).
6. Las covariables deben poseer una relación lineal con la variable de respuesta. De no ser así, se debe aplicar una transformación para convertir la relación en lineal. Esta transformación puede ser logarítmica, recíproca, arcoseno, exponencial o de raíz.
*Nota: la prueba de los supuestos debe realizarse antes de hacer el ANCOVA propiamente dicho, ya que el incumplimiento de alguno de los supuestos invalida en análisis de la covarianza.

El primer supuesto sobre la relación lineal entre la covariable y el desenlace se prueba haciendo un análisis de regresión lineal simple entre la covariable (X) y la respuesta (Y). La regresión debe ser de $p < 0.05$ (figura 1). El segundo supuesto se evalúa al probar si hay interacción entre la maniobra y la covariable (figura 2). El tercer supuesto de la independencia entre la covariable y la maniobra se prueba realizando un ANOVA, en el que la covariable (X) sea declarada como variable dependiente y los tratamientos sean declarados como variables independientes. La prueba de F debe resultar no significativa (figura 3).

Es importante tomar en cuenta algunas consideraciones: las covariables no deben estar relacionadas con las maniobras o condiciones experimentales y las covariables deben estar relacionadas con el desenlace (variable dependiente o de respuesta) para hacer un uso adecuado de ellas (cuadro I).

Una vez que comprobamos que se cumplen los supuestos de la prueba, podemos proceder a realizar el ANCOVA. Continuando con el ejemplo anterior, quisiéramos saber: 1) la media del peso del recién nacido ajustada por la ganancia de peso gestacional para las mujeres con anemia y sin anemia, 2) si el nivel de hierro (anemia frente a no anemia), es un predictor del peso del neonato y 3) si es diferente la media de peso de los recién nacidos entre las mujeres con anemia y sin anemia, independientemente de la ganancia de peso en el embarazo (ajustada por ganancia de peso).

En la figura 4 se pueden observar los pasos que se deben seguir para realizar el análisis en el programa SPSS y los resultados obtenidos a partir de este ejemplo. La media de peso sin ajustar de los recién nacidos con madre sin anemia fue de 3175 g y de 2649 g para los que tuvieron una madre con anemia. Cuando se asignó la misma ganancia de peso a todas las participantes, que en este caso fue de 11.55 kg, que fue la media de ganancia de peso para toda la muestra, se observó que el promedio ajustado de peso del recién nacido con madre sin anemia fue de 3104 g, mientras que la media ajustada para neonatos de madre con anemia fue de 2720 g. El estado de hierro resultó ser un predictor significativo del peso al nacer. En el cuadro en el que se presentan los estimadores de parámetro (*parameter estimates*) se observa un valor de $p = 0.009$ para el estado de hierro. También se observa que la ganancia de peso es un predictor significativo del peso al nacer ($p = 0.002$).

Dado que la ganancia de peso es un predictor significativo, resulta importante evaluar el efecto del estado de anemia eliminando el efecto de la ganancia de peso de la madre durante el embarazo. Por último, se puede observar que la diferencia de medias de peso del recién nacido entre los grupos fue de 384 g (*pairwise comparisons*) y que dicha diferencia fue estadísticamente significativa ($p = 0.009$) —por medio de la prueba de hipótesis de F (*univariate test*)—. En este caso podríamos haber ingresado como efectos aleatorios los estados o las colonias de los distintos hospitales donde se atendieron las mujeres embarazadas.

Resumiendo, el ANCOVA es una prueba que utiliza el análisis de varianza y el de regresión para eliminar la variabilidad que existe entre un factor de posible confusión que no pudo ser controlado en el estudio (llamado covariable) y proporciona medias ajustadas del desenlace para cada una de las maniobras para estimar el efecto de la maniobra sobre el desenlace sin el efecto de una covariable.

Declaración de conflicto de interés: los autores han completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no fue reportado alguno en relación con este artículo.

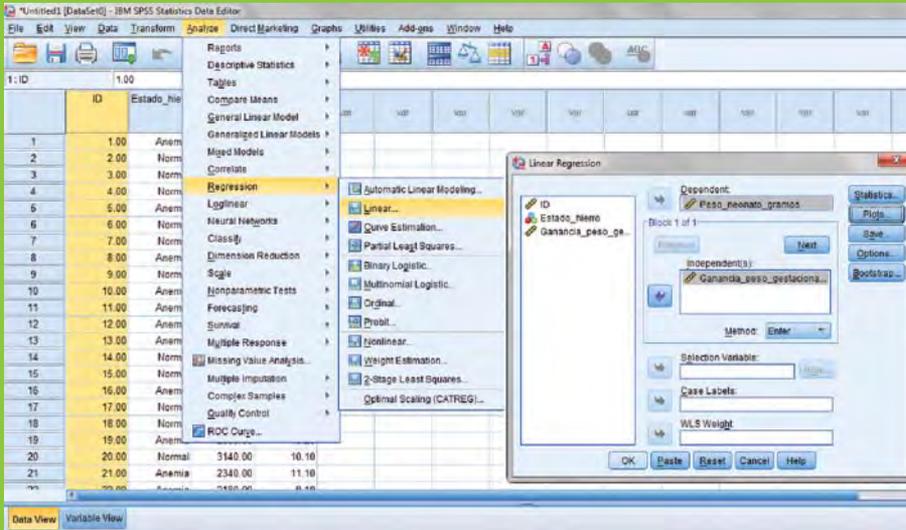


Figura 1 Comprobación del supuesto de relación lineal entre la covariable y el desenlace



Variables Entered/Removed^a

Model	Variables Entered
1	ganancia_peso_gestacional_kg ^b

a. Dependent Variable: peso_neonato_gramos
b. All requested variables entered

ANOVA^a

	Model	Sum of Squares	df	Mean Square	F	Sig.
	1 Regression	3062073.969	1	3062073.969	14.331	.001 ^b
	Residual	8119623.531	38	213674.303		
	Total	11181697.50	39			

a. Dependent Variable: Peso_neonato_gramos
b. Predictors (Constant), Ganancia_peso_gestacional_kg

Model Summary

Model	R	R Square	Adjusted R Square
1	.523 ^a	.274	.255

a. Predictors: (Constant), Ganancia_peso_gestacional_kg

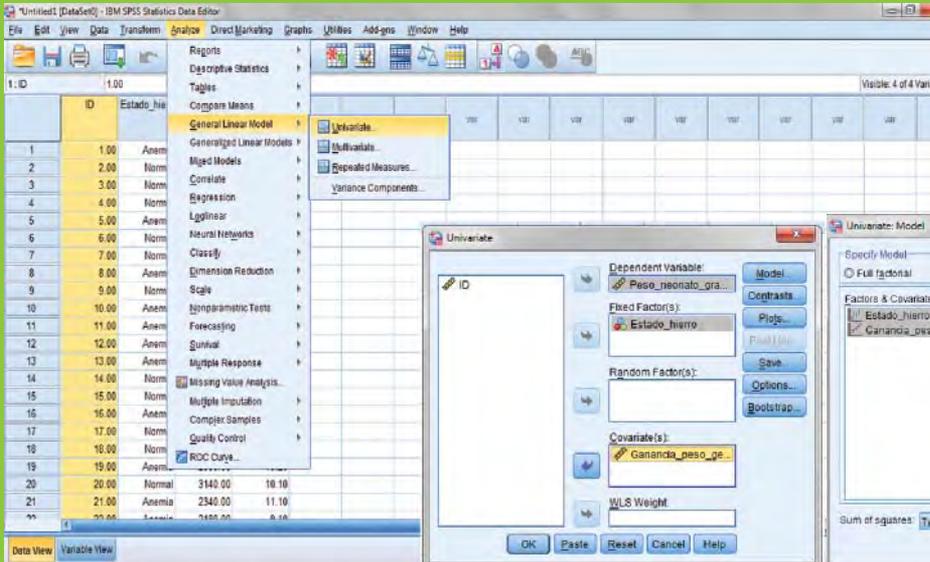


Figura 2 Comprobación del supuesto de homogeneidad de las pendientes



Tests of Between-Subjects Effects

Dependent Variable: Peso_neonato_gramos

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	5238956.76 ^a	2	2619478.379	16.309	.000
Intercept	3553151.452	1	3553151.452	22.122	.000
Estado_hierro*	5238956.759	2	2619478.379	16.309	.751
Ganancia_peso_gestacional_kg					
Error	5942740.741	37	160614.615		
Total	350429700.0	40			
Corrected Total	11181697.50	39			

a. R Squared = .469 (Adjusted R Squared = .440)

Between-Subjects Factors

	Value Label	N
Estado_hierro	.00	20
	1.00	20
	Normal	
	Anemia	



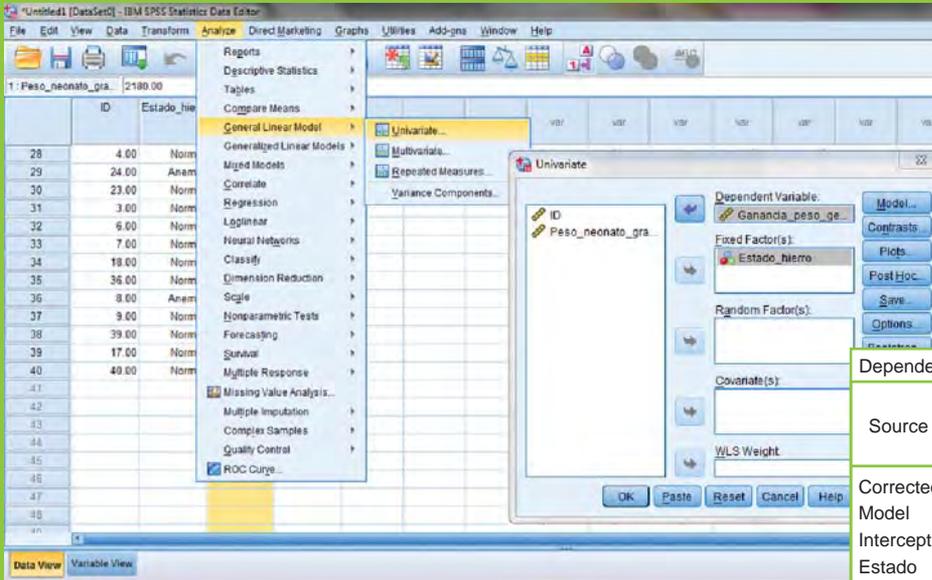


Figura 3 Comprobación del supuesto de independencia entre la covariable y la maniobra

Tests of Between Subjects Effects

Dependent Variable: Ganancia_peso_gestacional_kg

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	15.876 ^a	1	15.876	3.956	.054
Intercept	5336.100	1	5336.100	1329.791	.000
Estado_hierro*	15.876	1	15.876	3.956	.054
Error	152.484	38	4.013		
Total	5504.460	40			
Corrected Total	168.360	39			

a. R Squared = .094 (Adjusted R Squared = .070)

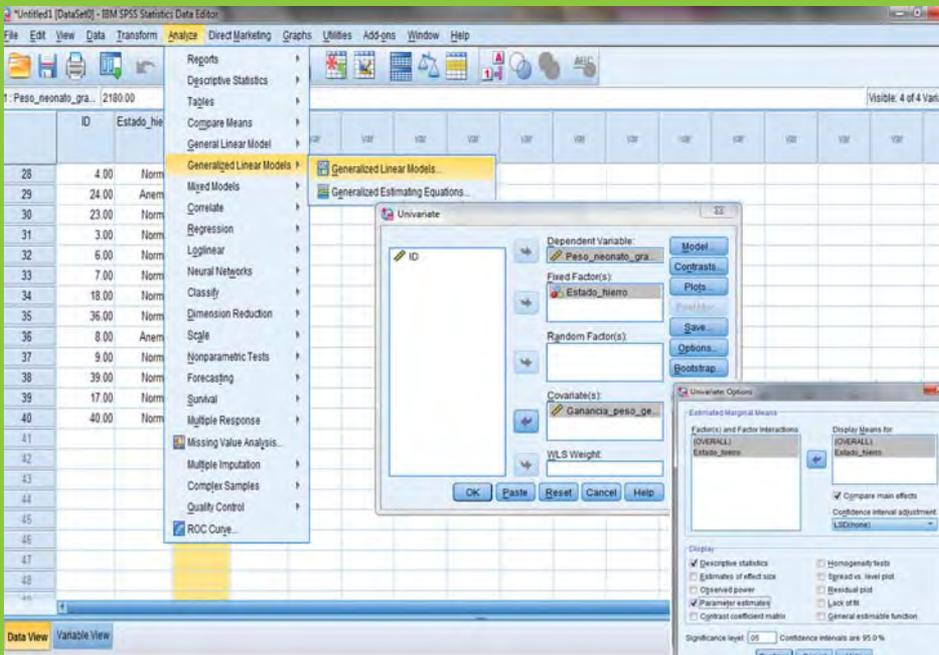


Figura 4 Procedimientos y resultados del análisis de la covarianza (ANCOVA)

Descriptive Statistics

Dependent Variable: Peso_neonato_gramos

Estado hierro	Mean	Std. Deviation	N
Normal	3175.5000	540.20927	20
Anemia	2649.0000	388.31485	20
Total	2912.2500	535.45326	40

Estimates

Dependent Variable: Peso_neonato_gramos				
(I) Estado hierro	Mean	Std. Error	95 % Confidence Interval	
			Lower Bound	Upper Bound
Normal	3104.244 ^a	95.827	2910.081	3298.408
Anemia	2720.256 ^a	95.827	2526.092	2914.419

a. Covariates appearing in the model are evaluated at the following values:
 Ganancia_peso_gestacional_kg = 11.5500.

Parameter Estimates

Dependent Variable: Peso_neonato_gramos						
Parameter	B	Std. Error	t	Sig.	95 % Confidence Interval	
					Lower Bound	Upper Bound
Intercept	1413.901	381.110	3.710	.001	641.698	2186.103
Ganancia_peso_gestacional_kg	113.104	33.835	3.343	.002	44.548	181.661
[Estado_hierro = .00]	383.989	138.832	2.766	.009	102.6888	665.289
[Estado_hierro = 1.00]	0 ^a					

a. This parameter is set to zero because it is redundant

Pairwise Comparisons

Dependent Variable: Peso_neonato_gramos						
(I) Estado hierro	(J) Estado hierro	Mean Difference (I-J)	Std. Error	Sig. ^b	95 % Confidence Interval	
					Lower Bound	Upper Bound
Normal	Anemia	383.989*	138.832	.009	102.688	665.289
Anemia	Normal	-383.989*	138.832	.009	-665.289	-102.688

Based on estimated marginal means
 * The mean difference is significant at the .05 level
 b. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Univariate Tests

Dependent Variable: Peso_neonato_gramos					
	Sum of Squares	df	Mean Square	F	Sig.
Contrast	1335432.395	1	1335432.395	7.650	.009
Error	6459009.936	37	174567.836		

The F tests the effect of Estado_hierro. This test is based on the linearly independent pairwise comparisons among the estimated marginal means.

^aCentro de Adiestramiento en Investigación Clínica (CAIC), Coordinación de Investigación en Salud, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social
^bInstituto Nacional de Psiquiatría "Ramón de la Fuente Muñiz", Secretaría de Salud

Distrito Federal, México
 Comunicación con: Juan O. Talavera
 Correo electrónico: jotalaverap@uaemex.mx

Bibliografía

1. Fisher RA. Statistical methods for research workers. Tenth edition. New York: Hafner; 1948.
2. Fisher RA. The analysis of covariance method for the relation between a part and the whole. Biometrics. 1947;3(2):65-8.
3. Porter AC, Raudenbush SW. Analysis of covariance: Its model and use in psychological research. J Couns Psychol. 1987;34(4):383-92.



Investigación clínica XVII

Prueba χ^2 , de lo esperado a lo observado

Rodolfo Rivas-Ruiz,^a Osvaldo D. Castelán-Martínez,^a Marcela Pérez,^a Juan O. Talavera^a

Clinical research XVII. χ^2 test, from the expected to the observed

When you want to show if there is a statistical association or differences between categorical variables, it is recommended to use the χ^2 test. This nonparametric test is one of the most used in clinical research; it contrasts nominal or ordinal qualitative variables that are observed in clinical practice. This test calculates the p value that determines whether differences between groups are real or due to chance. The χ^2 test is the basis of other tests to analyze qualitative ordinal variables as χ^2 for linear trend, which compares three groups with two outcomes or McNemar test which contrasts two related samples (a before and afterward comparison) or Mantel-Haenszel χ^2 , which controls for potential confounding variables. When using small samples, where the expected is less than 5, Fisher's exact test should be used. These tests are the most widely used in the medical literature, however, do not give us the magnitude or the direction of the event and a proper interpretation that requires clinical judgment is needed.

Key words

chi-square
qualitative variables
statistics

En la parte XVI de esta serie mostramos que para analizar variables cuantitativas de libre distribución se utiliza la U de Mann-Whitney. En este artículo, al igual que en el anterior, presentamos cómo se estudian las variables categóricas (cualitativas) de libre distribución, ya sea ordinales o nominales. Este tipo de pruebas están incluidas en la llamada *estadística no paramétrica*. De este grupo, la más característica es la χ^2 de Pearson, la cual contrasta la asociación entre dos grupos cuyo resultado es una variable dicotómica. Casi todas las investigaciones médicas utilizan esta prueba estadística, por el tipo de variable que contrasta. En la mayoría de los artículos, en el cuadro estadístico en el que se muestra que el estado basal de los sujetos es similar, se emplea la prueba de χ^2 para las variables categóricas. También se utiliza en los cuadros estadísticos de los resultados del estudio, para mostrar las diferencias entre los grupos analizados.

Esta prueba se basa en la diferencia entre lo observado en un experimento y lo que podría suceder por simple azar. La fórmula de χ^2 expresa la sumatoria de lo observado con lo esperado:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Donde:

O = eventos observados en el experimento.

E = eventos esperados por el azar.

Antes de iniciar el uso de χ^2 deben considerarse dos requisitos indispensables:

- Las frecuencias que se usan en este estadístico representan cuentas individuales y no rangos ni medidas de resumen.
- Las categorías son exhaustivas y mutuamente excluyente entre ellas, por lo tanto, cada individuo puede ser asignado a una categoría, pero únicamente a una.

Con esto en mente, pensemos en el siguiente ejemplo: si lanzamos 100 veces una moneda al aire, la hipótesis nula o de igualdad de proporciones (dado que en un principio debe asumirse que la moneda no tiene trampa) es que 50 veces caerá un lado de la moneda hacia arriba (cara) y 50 veces, el otro (cruz). El resultado es que el lado de la cara cae 47 veces y el lado de la cruz, 53 veces. Por lo tanto, las opciones posibles son dos: existe truco o el resultado se debe al azar. Para verificar qué es lo que sucede, realizamos la prueba de χ^2 en la que sustituimos de la siguiente forma:

Cuando se busca demostrar que existe relación o diferencias estadísticamente significativas, entre las variables categóricas se utiliza la prueba de χ^2 . Esta prueba no paramétrica es una de las más usadas en la investigación clínica, ya que contrasta variables cualitativas nominales u ordinales observadas en la práctica clínica. Con esta prueba se calcula el valor de p para decidir si las diferencias entre los grupos son reales o se deben al azar. Esta prueba genéricamente recibe el nombre de χ^2 o chi cuadrada (dependiendo de las preferencias idiomáticas) y es la base de otras pruebas con las que se analizan las variables cualitativas ordinales: la χ^2 de tendencia lineal, con la que se comparan tres grupos que pueden tener dos desenlaces; la prueba de McNemar, que sirve para

contrastar dos muestras relacionadas (antes y después de una maniobra); o la χ^2 de Mantel-Haenszel, que controla el efecto de variables potencialmente confusoras. Para muestras pequeñas, en las que el resultado esperado es menor de 5, se debe utilizar la prueba exacta de Fisher. Como se puede ver, este grupo de pruebas es el de mayor uso en los artículos médicos, sin embargo, no dan cuenta de la magnitud ni de la dirección del evento, por lo que su adecuada interpretación requiere del juicio clínico.

Resumen

Palabras clave
chi-cuadrada
variables cualitativas
estadística

$$Cara = (O - E)^2/E = (47 - 50)^2/50 = (3)^2/(50) = 0.18$$

$$Cruz = (O - E)^2/E = (53 - 50)^2/50 = (3)^2/(50) = 0.18$$

Finalmente se realiza la sumatoria (Σ) de estas dos posibilidades (0.18 + 0.18), cuyo resultado es igual a 0.36, que corresponde al valor de χ^2 .

Para determinar si es o no significativo, este valor se busca en la tabla de valores críticos de χ^2 (también llamada de contingencias), en el valor de p con

1 grado de libertad (gl): 0.36 se encuentra entre los valores críticos $p = 0.9$ y $p = 0.1$ y como está más cerca del primero se asume que le corresponde este. Dado que el valor de alfa (con significación) por lo general es ≤ 0.05 , entonces el valor de χ^2 (al que le corresponde una $p = 0.9$) no es significativo (figura 1).

El modelo de χ^2 tiene múltiples usos y particularidades, no solo funciona con variables cualitativas para dos grupos, si no que puede ser usado por tres grupos y con muestras pequeñas.

gl	0.995	0.99	0.975	0.95	0.9	0.1	0.05	0.025	0.01	0.005
1	0.0	0	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.01	0.02	0.051	0.103	0.216	4.605	5.991	7.378	9.21	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.948	11.45	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.86
5	0.412	0.554	0.831	1.145	1.61	9.236	11.071	12.833	15.086	16.75



Localizamos el valor de p que le corresponde a nuestra χ^2

Figura 1 Tabla de valores críticos de χ^2

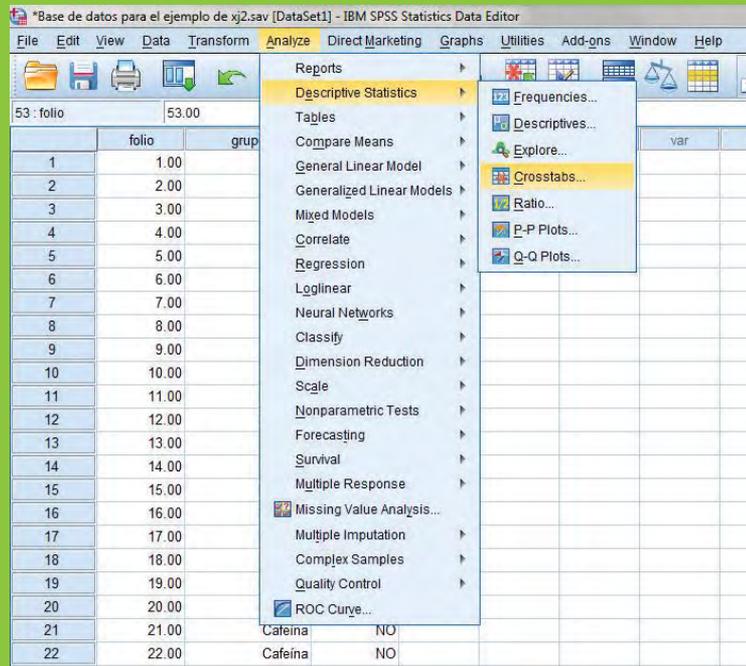
Ejemplo de un análisis con el programa estadístico SPSS

Se desea determinar si la aplicación de cafeína comparada con la de teofilina es eficaz para evitar una nueva intubación en los recién nacidos prematuros con dificultad para respirar. Se diseña un ensayo clínico aleatorizado con dos grupos:

- Uno al que se le administra cafeína.
- Otro al que se le administra teofilina.

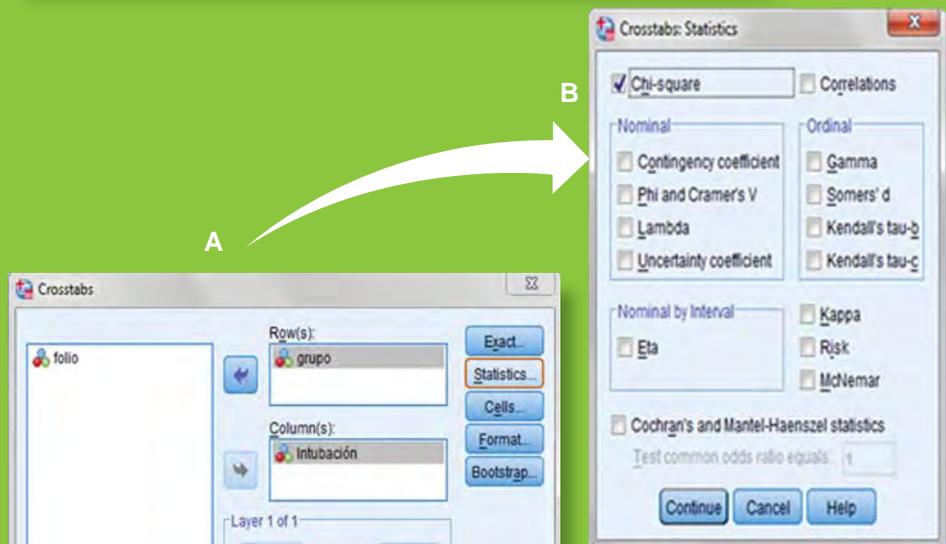
1

En el menú “Analizar” se selecciona la opción “estadísticos descriptivos” y en ésta, las tablas de 2 x 2 (*crosstabs*).

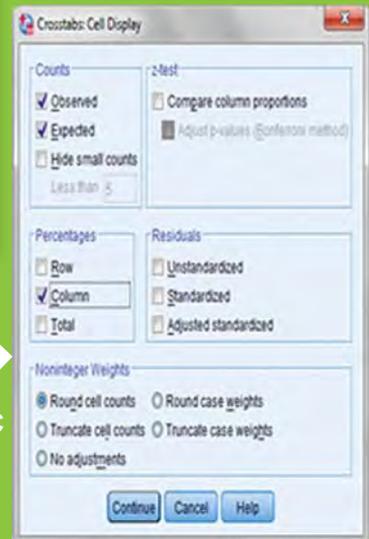
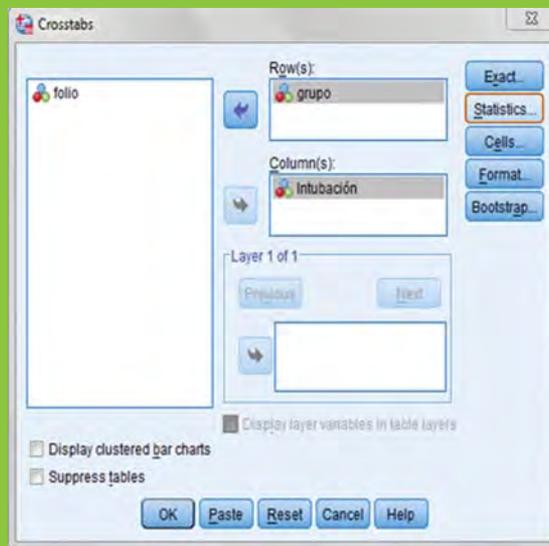


2

Al desplegar el menú de *crosstabs* (A), se selecciona la prueba estadísticas y después la prueba de χ^2 (B).



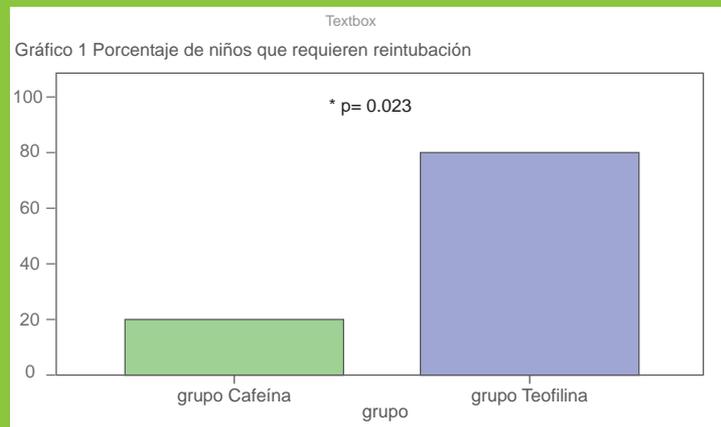
Una buena opción es seleccionar celdas o *cells* en el menú de *crosstabs* (A) y al desplegarse (C), activar las casillas de los valores observados y los esperados, estos últimos con la finalidad de seleccionar la prueba apropiada, ya que si en alguna de las celdas el valor esperado es menor de 5 se debe seleccionar la prueba exacta de Fisher (el programa la calcula automáticamente al seleccionar la prueba de χ^2).



Finalmente, una manera de saber la magnitud y dirección de esta comparación es elegir los porcentajes de respuesta por columna, de esta manera será posible determinar qué tanto funciona la cafeína en los niños prematuros al compararla con la teofilina.

3

En una tabla de 2×2 se puede ver el resultado de comparar la proporción de reintubación en el grupo con cafeína y en el grupo con teofilina. Por convención, en las columnas se tabula la variable dependiente, en este caso, la reintubación.



		Reintubación		Total
		Sí	No	
Cafeína	Recuento	2	38	40
	Frecuencia esperada	5.5	34.5	40.0
	% dentro de reintubación	18.2	55.1	50.0
Teofilina	Recuento	9	31	40
	Frecuencia esperada	5.5	34.5	40.0
	% dentro de reintubación	81.8	44.9	50.0
Total	Recuento	11	69	80
	Frecuencia esperada	11.0	69.0	80.0
	% dentro de reintubación	100.0	100.0	100.0

4

Abajo de la tabla de 2×2 aparecen las pruebas de χ^2 . La dificultad que tiene esta salida es la correcta identificación del valor p correspondiente. Debido a que nuestro ejemplo se trata de una comparación de grupos independientes y en las celdas el valor esperado más pequeño es de 5.5, se usa la prueba de χ^2 de Pearson ($p = 0.023$). Si en alguna celda hubiera un valor < 5 se usaría el estadístico exacto de Fisher con significación bilateral ($p = 0.048$). Si hubiera tres grupos (si quisiéramos comparar un grupo con cafeína, otro con teofilina y un grupo placebo), se usaría la asociación lineal por lineal ($p = 0.024$).

Pruebas de chi-cuadrado					
	Valor	gl	Sig. asintótica (bilateral)	Sig. exacta (bilateral)	Sig. exacta (unilateral)
Chi cuadrado de Pearson	5.165 ^a	1	0.023		
Corrección por continuidad ^b	3.794	1	0.051		
Razón de verosimilitudes	5.530	1	0.019		
Estadístico exacto de Fisher				0.048	0.024
Asociación lineal por lineal	5.100	1	0.024		
N de casos válidos	80				

^a0 casillas (0.0 %) tienen una frecuencia esperada inferior a 5. La frecuencia mínima esperada es 5,50

^bCalculado solo para una tabla de 2×2

Para conocer la magnitud, se debe ver la tabla de 2×2 (**3**), en la parte donde se calculan los porcentajes. En este ejemplo, el grupo de niños tratados con cafeína tuvo una incidencia de reintubación de 18.2 %, mientras que la incidencia de intubación del grupo de teofilina fue de 81.8 %. La prueba de χ^2 muestra que la diferencia entre estos dos grupos es estadísticamente significativa. ($p = 0.023$) (**4**).

Prueba exacta de Fisher

Es una prueba estadística alternativa a la χ^2 para comparar dos proporciones binomiales independientes en una muestra pequeña, en la que en alguna de las casillas el valor esperado es < 5 . Calcula las probabilidades exactas de obtener un resultado determinado cuando la hipótesis nula es cierta. Es muy útil cuando la aproximación normal a la distribución binomial podría no ser aplicable, por ejemplo, cuando se presentan pocas observaciones en una de las casillas de la tabla de 2×2 o en proporciones extremas.

Prueba de McNemar (antes y después)

La prueba de McNemar es un caso especial de la prueba binomial que se usa para comparar dos proporciones de muestras pareadas, por ejemplo en un estudio de antes y después en el que un paciente o un grupo es su propio control. Con frecuencia se emplea en los ensayos clínicos cuyos resultados son variables dicotómicas que se registran dos veces para cada paciente bajo diferentes condiciones y se busca evaluar los cambios ocurridos al transcurrir un periodo. Las condiciones pueden representar diferentes tratamientos o diferentes tiempos de medición. El objetivo es comparar las tasas de respuesta en los dos conjuntos de condiciones. Debido a que las mediciones provienen de los mismos pacientes, no es posible asumir que los grupos sean independientes, por tal motivo no es posible utilizar la χ^2 o la prueba exacta de Fisher.

Estudios típicos, más específicos, en los que se aplica la prueba de McNemar son los que comparan las tasas de respuesta de dos tratamientos oftálmicos cuando ambos se proporcionan a cada paciente, uno en cada ojo.

χ^2 de Mantel-Haenszel

Cuando se reconoce que la asociación observada puede ser influida por una variable externa es pertinente utilizar la χ^2 de Mantel-Haenszel (variante de la χ^2 de Pearson), cuya función es controlar el efecto de una variable potencialmente confusora sobre la variable de resultado, por ejemplo: al estudiar la asociación entre el consumo de café y el desarrollo de gastritis, se

identifica que el consumo de café está asociado con el hábito tabáquico, que, a su vez, se asocia con el riesgo de gastritis, por lo tanto, se debe investigar si la asociación entre consumo de café y gastritis es independiente del hábito tabáquico. Es importante calcular el estimador de riesgo (*odds ratio* o riesgo relativo) de gastritis, después calcular el estimador de riesgo por estratos de la variable potencialmente confusora, es decir, el riesgo de gastritis en quienes fuman y en quienes no fuman. Después se calcula el estimador de riesgo ajustado de Mantel-Haenszel para el análisis estratificado, con base en los riesgos interestrato. Para saber si la asociación entre consumo de café y la gastritis es independiente del hábito tabáquico se calcula la χ^2 de Mantel-Haenszel. El valor que arroja esta prueba se busca en la tablas de χ^2 : si es mayor al valor crítico para un grado de libertad, la asociación es estadísticamente significativa independientemente del hábito tabáquico.

χ^2 de tendencia lineal

También denominada asociación lineal por lineal, es la prueba más apropiada para analizar situaciones en las que una variable de exposición ordenada tiene tres o más categorías (como en el ejemplo anterior) y la variable de resultado es de tipo binario.

Conclusiones

La χ^2 es la base de la mayoría de las pruebas estadísticas que se usan en la práctica clínica. Se basa en la diferencia de lo observado menos lo esperado, fenómeno que más se asocia con la práctica clínica, en la que el resultado es si un paciente mejora o no o si una maniobra previene o no una condición. Esta prueba solo ayuda a saber si el resultado se debe a la maniobra y no al azar. Para determinar la magnitud del evento se requiere observar los porcentajes de respuesta por grupo, graficar esta o aplicar otras pruebas de asociación, como la razón de momios y el riesgo relativo.

Declaración de conflicto de interés: los autores ha completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no fue reportado alguno en relación con este artículo.

^aCentro de Adiestramiento en Investigación Clínica, Coordinación de Investigación en Salud, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Distrito Federal, México

Comunicación con: Rodolfo Rivas-Ruiz.
Correo electrónico: rivasrodolfo@gmail.com

Bibliografía

1. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst.* 1959;22(4):719-48.
2. Henderson-Smart DJ, Davis PG. Prophylactic methylxanthines for endotracheal extubation in preterm infants. *Cochrane Database Syst Rev.* 2010;(12):CD000139.
3. Portney LG, Watkins MP. *Foundations of clinical research: applications to practice.* Third edition. New Jersey: Pearson/Prentice Hall; 2009.
4. Armitage P, Berry G, Matthews JNS. *Statistical methods in medical research.* Fourth edition. Massachusetts: Blackwell Science; 2002.
5. Peat JK, Barton B. *Medical statistics: a guide to data analysis and critical appraisal.* Massachusetts: Blackwell 2005.
6. Le Chap T. *Introductory biostatistics.* New Jersey: Wiley-Interscience; 2003.
7. Bland M. *Introduction to medical statistics.* Third edition. US: Oxford University Press; 2003.
8. Dawson B, Trapp R. *Basic and clinical biostatistics.* Fourth edition. US: Lange Medical Books-McGrawHill; 2004.
9. Talavera JO, Wacher-Rodarte NH, Rivas-Ruiz R. Clinical research III. The causality studies. *Rev Med Inst Mex Seguro Soc.* 2011;49(3):289-94. Texto libre en http://revistamedica.imss.gob.mx/index.php?option=com_multicategories&view=article&id=1410:investigacion-clinica-iii-estudios-de-causalidad&catid=501:temas-de-actualidad&Itemid=664
10. Talavera JO, Rivas-Ruiz R. Clinical research IV. Relevancy of the statistical test chosen. *Rev Med Inst Mex Seguro Soc.* 2011;49(4):401-5. Texto libre en http://revistamedica.imss.gob.mx/index.php?option=com_content&view=article&id=1432:investigacion-clinica-iv-pertinencia-de-la-prueba-estadistica&catid=139:2011&Itemid=511
11. TalaveraJO,Rivas-RuizR. Clinicalresearch V. Sample size. *Rev Med Inst Mex Seguro Soc.* 2011;49(5):517-22. Texto libre en http://revistamedica.imss.gob.mx/index.php?option=com_content&view=article&id=1452:investigacion-clinica-v-tamano-de-muestra&catid=139:2011&Itemid=511
12. Rivas-Ruiz R, Pérez-Rodríguez M, Talavera JO. Del juicio clínico al modelo estadístico. Diferencia de medias. Prueba *t* de Student. *Rev Med Inst Mex Seguro Soc.* 2013;51(3):300-3. Texto libre en http://revistamedica.imss.gob.mx/index.php?option=com_multicategories&view=article&id=2077:51-3-11-investigacion&catid=785:temas-de-actualidad&Itemid=775
13. Rivas-Ruiz R, Moreno-Palacios J, Talavera JO. Investigación clínica XVI. Diferencia de medianas con la U de Mann-Whitney. *Rev Med Inst Mex Seguro Soc* 2013;51(4):414-9.

Investigación clínica VI.
Relevancia clínica

Centro de Adiestramiento en Investigación Clínica,
Coordinación de Investigación en Salud, Centro Médico Nacional Siglo XXI,
Instituto Mexicano del Seguro Social, Distrito Federal, México

Comunicación con: Juan O. Talavera.
Correo electrónico: jotalaverap@uaemex.mx

Resumen

En la clínica suele seleccionarse la maniobra que logra un desenlace cuando menos de 10 % de superioridad directa o cuando el número necesario a tratar es ≈ 10 . Si bien estos parámetros sirven para estimar la magnitud de una asociación, estamos obligados a diferenciar las medidas de impacto (riesgo atribuible, fracción prevenible), de asociación (riesgo relativo, razón de momios, razón de riesgo) y de frecuencia (incidencia y prevalencia), aplicables cuando el desenlace es nominal. También debemos identificar la forma de medir la fuerza de la asociación y la magnitud de la asociación cuando la variable de desenlace es cuantitativa. No es raro interpretar las medidas de asociación como si fueran de impacto, por ejemplo, para un riesgo relativo de 0.68 se asume una reducción de 32 % del desenlace sin considerar que ésta es una reducción relativa que igual proviene de una relación de 0.4/0.6, 0.04/0.06 o 0.00004/0.00006; sin embargo, la reducción directa es de 20 % (60-40 %), de 2 % y de 2 por cada 100 000, respectivamente. Por lo tanto, para estimar el impacto de una maniobra es importante contar con la diferencia directa o el número necesario a tratar.

Palabras clave

medidas de asociación, exposición, riesgo o desenlace
riesgo relativo

Summary

Usually, in clinical practice the maneuver selected is the one that achieve a favorable outcome with a direct percentage of superiority of at least 10 %, or when the number needed to treat ≈ 10 . While this percentage difference is practical to estimate the magnitude of an association we need to differentiate the impact measures (attributable risk, preventable fraction), measures of association (RR, OR, HR) and frequency measures (incidence and prevalence), applicable when the outcome is nominal. And we must identify ways to measure the strength of association and the magnitude of the association when the outcome variable is quantitative. It is not uncommon to interpret the measures of association as if they were impact measures. For example, for a RR of 0.68, its common to assume a 32 % reduction of the outcome, but, we must consider that this is a relative reduction, which comes from relation of 0.4/0.6, 0.04/0.06, or 0.00004/0.00006, however the direct reduction is 20 % (60 % - 40 %), 2 %, and 2 per 100 000 respectively. Therefore, to estimate the impact of a maneuver it is important to have the direct difference and / or NNT.

Key words

measures of association, exposure, risk or outcome
relative risk

Introducción

Aun teniendo un estudio bien diseñado, con un análisis estadístico y tamaño de muestra adecuados, donde se muestra significancia estadística en la asociación entre una maniobra y un desenlace (trátase de la asociación entre un factor de riesgo o maniobra preventiva y la ocurrencia de una enfermedad, o entre un factor pronóstico o maniobra terapéutica y la evolución de la enfermedad), el clínico tiene la necesidad de identificar la magnitud de dicha asociación

—impacto de la maniobra—, para considerar su utilidad en la práctica clínica cotidiana, donde las más de las veces se considera el beneficio de una maniobra terapéutica y suele seleccionarse la que logra un desenlace favorable con al menos 10 % de superioridad directa sobre otras. Esto significa que, por ejemplo, si el desenlace es supervivencia y la maniobra seleccionada es la *A*, se espera que ésta sea 10 % superior a la maniobra estándar *B* (70 % de supervivencia a dos años con la maniobra *A* contra 60 % de la maniobra *B*), o si el desenlace es el nivel de glucosa, entonces se espera

una reducción mínima de 10 % (de 140 a 126 mg/dL). Y si el desenlace es insuficiencia cardiaca, se espera una reducción mínima de 10 % del grado de falla cardiaca (que en su conjunto cuando menos exista 10 % más de pacientes que mejoraron su grado de falla cardiaca). Hay que hacer notar que la sustracción de una proporción a partir de otra se realizó en forma directa, mientras que para datos cuantitativos se estima 10 % a partir del valor de referencia.

En salud pública o medicina preventiva, las diferencias directas menores a 10 %, incluso de 4 a 7 %, resultan de suma relevancia dado que las poblaciones susceptibles pueden ser de millones de sujetos. Lo mismo sucede en la atención clínica donde el desenlace no deseado es de alrededor de 10 %, para lo cual cualquier reducción esperada será inferior al mismo y cuya relevancia dependerá de la gravedad y costo de la alteración. Por otro lado, en caso de eventos adversos diferencias incluso menores a 10 % resultan significativas, sobre todo dependiendo de la intensidad del mismo. No obstante, para la mayoría de las situaciones en la clínica se considera deseable una ganancia mínima de 10 %.

Si bien para los médicos clínicos resulta común y entendible una diferencia de porcentaje para estimar el impacto de una asociación, en la literatura existe una serie de cálculos conocidos como *medidas de impacto*, que si bien resultan discretamente más elaborados acaban siendo una asociación entre proporciones. En el proceso de obtener las medidas de impacto se estiman las *medidas de asociación*, indicadores que evalúan la fuerza con la que una variable o característica se asocia con otra, mismas que no tendrían sentido si no se acompañaran de la certeza de que dicha asociación es real y no debida al azar, para lo cual se estima la *significancia estadística* (una asociación es real cuando el valor de $p < 0.05$). Antes que estos dos tipos de medidas, durante el proceso de manejo de datos tenemos que hacer uso de lo que se conoce como *medidas de frecuencia*, las cuales miden el número absoluto de eventos. Debe resaltarse que en la mayoría de las ocasiones lo que observamos en los artículos son *medidas de frecuencia relativa*, donde se relaciona el número de eventos con el número total de individuos en la población o muestra en estudio, de tal forma que en una etapa posterior se pueden realizar comparaciones entre grupos con n distinta (cuadro I).

En la clínica, las mediciones de la asociación entre dos variables (maniobra y desenlace) a través de riesgo relativo (RR), razón de momios (OR, RM) y razón de riesgo (*Hazard risk ratio*, HR) son comunes y se interpretan en forma similar; variables con un valor inferior a 1 se consideran de protección mientras que aquellas con valor por arriba de 1, de riesgo. De esta forma tenemos que el riesgo común de la población o muestra de sufrir o tener el evento de interés sin identificar ningún factor ya sea de protección o de riesgo es de 1 (que corresponde a la incidencia o prevalencia del evento en el total de la muestra o población en estudio). Pero si identificamos un factor de riesgo vemos que la incidencia en

este subgrupo se incrementa y en el que no la tiene se disminuye en relación con el riesgo de la población o muestra total. Por ejemplo, si consideramos el uso de aspirina para prevenir infarto del miocardio en población donde la incidencia a un año es de 1.6 %, la incidencia en el grupo expuesto a aspirina será de 1.3 % mientras que en el grupo control de 1.9 % con un riesgo relativo de 0.68 (0.013/0.019), lo que significa que hay una reducción del riesgo relativo de 32 %. Hasta aquí parece haber una asociación entre el uso de la aspirina y la reducción de infarto, pero habrá que revisar el intervalo de confianza de 95 % para dicho riesgo relativo: si el intervalo en sus límites (inferior y superior) está por debajo de la unidad se considera estadísticamente significativo, pero si el valor superior rebasa la unidad (1) entonces no es estadísticamente significativo y, en consecuencia, no puede descartarse que el valor puntual observado de 0.68 se deba al azar. De igual manera, cuando hablamos de un factor de riesgo es de esperar que el límite inferior del intervalo de confianza de 95 % se encuentre por arriba de la unidad (1) para que sea estadísticamente significativo (cuadro II).

Las medidas de frecuencia, asociación e impacto están basadas en la presencia o no de un evento o desenlace y, por lo tanto, se trata de una variable nominal, pero en la clínica existen numerosas variables de desenlace que se miden a través del cambio en el valor de una variable cuantitativa, en la cual de igual forma interesa conocer la fuerza y magnitud de la asociación, por lo que resulta relevante contar con un equivalente.

En el cuadro III se muestran en un contexto global las medidas de frecuencia relativa, medidas de asociación y medidas de impacto, descritas básicamente para una variable dependiente nominal. Se agregan otras medidas también aplicables que pueden definir la fuerza de asociación —medidas de asociación—:

- El coeficiente de determinación r^2 , que mide el porcentaje de explicación de una variable a partir de la otra y que es el cuadrado de la r obtenida en una correlación, para este caso coeficiente phi.
- El coeficiente beta, valor obtenido durante un modelo de regresión (en este caso logística), que corresponde al logaritmo de la razón de momios.
- La R^2 similar a la r^2 , cuyo resultado se obtiene del modelo de regresión.

En cuanto a la magnitud de la asociación, del resultado de un modelo de regresión se puede obtener la probabilidad estimada de la ocurrencia de un fenómeno ($\hat{y} = 1 / (1 + e^{-(a + b1X1...)})$), que en la base de la ecuación para su cálculo suma los coeficientes beta de las distintas variables, y, finalmente, calcula su RM global. Con esta ecuación si se comparan dos tratamientos se puede estimar la diferencia de dicha probabilidad (diferencia de proporciones), aun ajustada por múltiples variables de interés; de igual forma se pueden comparar

Cuadro I | Tabla de doble entrada para medidas de frecuencia relativa (ejemplo), asociación e impacto

	Desenlace +	Desenlace –	Total
Expuestos (tratados)	a 5	b 95	a + b = 100
No expuestos (placebo)	c 15	d 85	c + d = 100
Total	a + c = 20	b + d = 180	

		Ejemplo	Interpretación
Incidencia de expuestos (Ie)	$Ie = a/a + b$	$5/100 = 0.05$ en 100 sujetos o 5 %	5 casos nuevos
Incidencia de observados o no expuestos (Io)	$Io = c/c + d$	$15/100 = 0.15$	15 casos nuevos en 100 sujetos o 15 %
Riesgo relativo (RR) o del riesgo. El riesgo está por debajo de la unidad	$RR = Ie/Io$	$0.05/0.15 = 0.33$ Reducción relativa	Existe una protección.
Reducción absoluta del RR (RAR) (riesgo atribuible [RA])	$RAR = Ie - Io$	$0.15 - 0.05 = -0.1$	La reducción directa del riesgo atribuida al tratamiento es de 10 %
Reducción del riesgo relativo (RRR) (fracción prevenible [FP])	$RRR = 1 - RR \times 100$	$1 - 0.33 \times 100 = 67 \%$	Hay reducción de 67 % (reducción relativa).

Prevalencia de los expuestos (Pe)	$Pe = a/a + b$		Número de eventos en el grupo expuesto (se usa en los estudios transversales)
Prevalencia de los no expuestos (Po) control (se usa en los estudios transversales)	$Po = c/c + d$	grupo no expuesto o	Número de eventos en el
Razón de momios (RM)	$a \times d/b \times c$	$RM = 5 \times 85/15 \times 95$ $RM = 425/1.425$ $RM = 0.29$ unidad	El grupo expuesto está protegido. El riesgo está por debajo de la

La incidencia y prevalencia son medidas de frecuencia, el riesgo relativo y la razón de momios se consideran medidas de asociación y la reducción absoluta del riesgo y la reducción del riesgo relativo, medidas de impacto. Otra medida de asociación es la razón de riesgo, obtenida en el análisis de supervivencia de riesgos proporcionales de Cox (Hazard risk ratio, HRR). También se puede estimar el riesgo atribuible a partir de RM ($RAR = RM - 1/RM$)

Cuadro II | Ejemplos de RR e intervalos de confianza de 95 %

Ejemplos de estudios	A		B		RR (IC 95 %)	RR (IC 95 %)
	Eventos	Total	Eventos	Total		
Aspirina (A) contra placebo	65	5000	95	5000	0.68 (0.50, 0.94)	
Consumo de café (A) contra placebo (B)	25	5003	24	5000	1.04 (0.60, 1.82)	
Con dislipidemia (A) contra sanos (B)	205	5000	115	5000	1.78 (1.42, 2.23)	

Las aspirina tiene una RRR de 32 %, estadísticamente significativa; la dislipidemia tiene un incremento del RR de 78 %, estadísticamente significativo. El consumo de café tiene un incremento relativo de 4 %, no estadísticamente significativo
 RR = riesgo relativo, IC 95 % = intervalo de confianza de 95 %, RRR = reducción del riesgo relativo

Cuadro III | Medidas de asociación y equivalentes para variables cuantitativas

Variable dependiente cualitativa (nominal)			Variable dependiente cuantitativa	
Medidas de frecuencia	Medidas de asociación	Medidas de Impacto	Fuerza de asociación	Magnitud de la asociación
Incidencia ■ Tasa de incidencia ■ Incidencia acumulada	RR (razón de incidencia acumulada) HR (<i>Hazzard risk ratio</i>)	Riesgo atribuible (fracción etiológica, RAR) RRR, FP (fracción prevenible)	r^2 R^2	% de diferencia de las medias Coeficiente b % de diferencia de las medias a través de la ecuación de la regresión ($\hat{y} = a + b1X1$)
Prevalencia ■ Prevalencia puntual ■ Prevalencia de Periodo	RM (razón de prevalencias o productos cruzados)			
	r Coeficiente b R^2	% de diferencias de proporciones a través la ecuación de probabilidad $\hat{y} = 1/1 + e^{-(a + b1X1...)}$		

$$\hat{y} = 1/1 + e^{-(a + b1X1...)} = \text{probabilidad del evento}$$

RRR = reducción del riesgo relativo

El NNT (número necesario a tratar) es una relativamente nueva forma de estimar la magnitud de la asociación

distintas probabilidades de ocurrencia de un fenómeno por exposición a distintos valores de una variable cuantitativa.

En el mismo cuadro III se muestra cuando la variable dependiente es cuantitativa: las unidades para medir la fuerza de la asociación se limitan a la r^2 de Pearson, al coeficiente b y a la R^2 , estas dos últimas como resultado del modelo de regresión lineal.

Finalmente, para evaluar la magnitud de la asociación de una variable cuantitativa se utilizan las diferencias de las medias, más específicamente la proporción de la diferencia de las medias, ya sea estimada en forma directa o como resultado

de la ecuación de regresión (en la regresión lineal el valor de la variable dependiente se obtiene en forma directa).

Una medida para la magnitud de la asociación que se ha popularizado en la actualidad es el *número necesario a tratar* (NNT = 1/RAR), el cual se refiere a la cantidad de sujetos que se requiere tratar para obtener el beneficio en uno; cuando este número resulta negativo se le conoce como *número necesario a dañar*. Por lo tanto, para definir si una maniobra es clínicamente significativa puede seguir usándose una diferencia directa de 10 % o el número necesario a tratar

(NNT), en el que si bien no hay un parámetro preestablecido se considera ideal un valor alrededor de 10, que correspondería tratar 10 sujetos para obtener en uno el beneficio deseado (equivalente a 10 %).

Comentarios

El uso apropiado de las medidas de frecuencia, asociación o impacto y sus equivalentes resultan indispensables para evitar errores comunes que se cometen en la práctica clínica. No es

raro interpretar las medidas de asociación como si fueran medidas de impacto, por ejemplo, si la OR o RM, el RR o el HR de una maniobra es de 0.68 se asume una reducción de 32 % del desenlace. Sin embargo, hay que tener en cuenta que ésta es una reducción relativa que igual proviene de una relación de 0.4/0.6, que de 0.04/0.06 o de 0.00004/0.00006; sin embargo, en el primer caso la reducción directa es de 20 % (60 a 40 %), en el segundo de 2 % y en el tercero de dos por cada 100 000. Por lo tanto, para estimar el impacto de una maniobra es importante contar con la diferencia directa o el NNT.

Bibliografía

1. Cordell WH. Number needed to treat (NNT). *Ann Emerg Med* 1999;33:433-436.
2. Feinstein AR. *Principles of medical statistics*. New York, NY: Chapman and Hall/CRC; 2002.
3. Guyatt GH, Sackett DL, Cook DJ. Users guides to the medical literature. II. How to use an article about therapy or prevention. B. What were the results and will they help in caring for my patients? Evidence Based Medicine Working Group. *JAMA* 1994;271:59-63.

Para el cálculo en línea

4. KT Clearing House. [Sitio web]. Odds ratio to NNT converter. Disponible en <http://ktclearinghouse.ca/cebm/practise/ca/calculators/ortonnt>
5. Sociedad Española de Hipertensión/Liga Española para la Lucha contra la Hipertensión Arterial. [Sitio web]. Odds ratio, riesgo relativo y número necesario a tratar. Disponible en <http://www.seh-lelha.org/oddsratio.htm>
6. University of British Columbia. [Sitio web]. UBC clinical significance calculator. Disponible en <http://spph.ubc.ca/sites/healthcare/files/calc/clinsig.html>



Investigación clínica XX

Del juicio clínico a la regresión logística múltiple

Ricardo Berea-Baltierra,^a Rodolfo Rivas-Ruiz,^b Marcela Pérez-Rodríguez,^b Lino Palacios-Cruz,^c Jorge Moreno,^d Juan O. Talavera^b

Clinical research XIX. From clinical judgment to multiple logistic regression model

The complexity of the causality phenomenon in clinical practice implies that the result of a maneuver is not solely caused by the maneuver, but by the interaction among the maneuver and other baseline factors or variables occurring during the maneuver. This requires methodological designs that allow the evaluation of these variables. When the outcome is a binary variable, we use the multiple logistic regression model (MLRM). This multivariate model is useful when we want to predict or explain, adjusting due to the effect of several risk factors, the effect of a maneuver or exposition over the outcome. In order to perform an MLRM, the outcome or dependent variable must be a binary variable and both categories must mutually exclude each other (i.e. live/death, healthy/ill); on the other hand, independent variables or risk factors may be either qualitative or quantitative. The effect measure obtained from this model is the odds ratio (OR) with 95 % confidence intervals (CI), from which we can estimate the proportion of the outcome's variability explained through the risk factors. For these reasons, the MLRM is used in clinical research, since one of the main objectives in clinical practice comprises the ability to predict or explain an event where different risk or prognostic factors are taken into account.

Keywords Palabras clave

Logistic models	Modelos logísticos
Causality	Causalidad
Biomedical research	Investigación biomédica

Desde que el ser humano tuvo conciencia de los peligros y oportunidades que a su alrededor podrían suceder, se originó la necesidad de determinar qué factores podrían estar relacionados y, finalmente, ser predictores o favorecedores; sobre todo, se interesó en aquellos que podrían ser controlados o modificados. Finalmente, en nuestra introspección y a través de nuestra capacidad de observación y de evaluación, los seres humanos nos dimos cuenta de que la mayor parte de los fenómenos que analizamos y muchas veces estudiamos están determinados por distintos factores distales y proximales, que en conjunto originan la presencia o ausencia de algún resultado.

Uno de los principales objetivos de la práctica clínica es predecir o explicar un evento en el que se tomen en cuenta diferentes factores de riesgo o pronósticos, de tal forma que nos permita estimar con un mayor grado de certeza un diagnóstico o un suceso a futuro, por ejemplo, al tratar de establecer la probabilidad de neumonía nosocomial en pacientes de reciente ingreso hospitalario a partir de múltiples variables, como antecedentes de diabetes mellitus, enfermedad pulmonar obstructiva crónica (EPOC), discapacidad funcional, edad del paciente, etcétera. Otro ejemplo de este carácter multifactorial en la predicción de un fenómeno consistiría en intentar predecir en un grupo de adolescentes con obesidad, a 10 años, cuáles son los factores más importantes (edad, actividad física, patrón de dieta, antecedentes hereditarios, etcétera) para el desarrollo de diabetes mellitus. En estos casos, si intentáramos aproximarnos matemáticamente, una alternativa adecuada podría ser aplicar el modelo de regresión logística múltiple.

En este modelo, la variable dependiente tiene dos valores posibles: la ausencia o la presencia de una característica (desarrollo o no de neumonía, diabetes presente o ausente). A su vez, las variables independientes o predictoras pueden ser continuas (edad), ordinales (estadio de Tanner, actividad física leve, moderada o intensa) o dicotómicas (discapacidad funcional, presente o ausente). Como ya lo mencionamos, en la vida diaria como en la práctica clínica, en pocas ocasiones existe una causa única para desarrollar una enfermedad o explicar un fenómeno; casi siempre coexisten múltiples factores de riesgo que hacen a un sujeto proclive a desarrollar o no una enfermedad.

Modelo de regresión logística múltiple

En el modelo de regresión logística múltiple se busca explicar o predecir la probabilidad de que ocurra o no un evento, el cual se identifica como la variable dependiente o Y ; se utiliza la ecuación de regresión, en la que se conoce a las variables X (X_1, X_2, \dots, X_k) como

Resumen

La complejidad del fenómeno de causalidad en la práctica clínica implica que el resultado de una maniobra no se deba únicamente a esta, sino a la interacción con otros factores del estado basal o variables que ocurran durante la maniobra. Esto requiere diseños metodológicos que permitan evaluar estas variables. Cuando el resultado es dicotómico, se usa la regresión logística múltiple (RLM). La RLM es un modelo multivariado útil cuando se requiere predecir o explicar, al ajustar por el efecto de distintos factores de riesgo, el efecto de una maniobra o exposición sobre el desenlace. Para realizar la RLM se requiere que el desenlace (o la variable dependiente) sea dicotómico

y mutuamente excluyente (por ejemplo, vivo/muerto, enfermo/sano); las variables independientes o factores de riesgo pueden ser cuantitativas o cualitativas. La asociación que se obtiene es la razón de probabilidades, también llamada razón de momios (RM), con intervalos de confianza (IC) del 95 % y con estas medidas se estima el porcentaje de la variabilidad del desenlace que se explica a partir de los factores de riesgo. Por estas razones, este modelo es el más usado en la investigación clínica, ya que uno de los principales objetivos de la práctica clínica es poder predecir o explicar un evento en el que se tomen en cuenta diferentes factores de riesgo.

variables independientes y se corresponden con las variables predictivas.

Este es un modelo multivariado muy popular y atractivo para resolver preguntas de investigación clínica. Una de sus características más importantes es que puede incluir variables predictivas de todos los tipos, es decir, continuas, ordinales o categóricas (figura 1).

Ahora bien, si en el análisis de nuestras variables de estudio decidimos utilizar la regresión logística múltiple, debemos tomar en cuenta los siguientes supuestos:

Las X_i son variables no aleatorias (fijas), es decir, son posibles predictoras de Y . La relación de las variables debe ser clara y permitir que se establezcan los factores de riesgo que puedan producir la enfermedad o el desenlace (Y). Es importante tener en cuenta que no deben utilizarse variables en los casos en los que no se conozca su relación con la enfermedad (Y).

La variable dependiente (Y) debe ser dicotómica y mutuamente excluyente, por ejemplo: neumonía (tiene o no tiene) o estado al final del tratamiento (vivo o muerto).

Los valores de X son independientes y no dependen de la presencia de Y ; por ejemplo, la presencia de neumonía nosocomial (Y) no es un factor de riesgo para el desarrollo de diabetes mellitus o de EPOC (X).

Una vez que conocemos que nuestra variable dependiente es dicotómica, se le codifica como 0 o 1. La probabilidad (*odd* en inglés) de que el evento suceda entre la probabilidad de que no suceda se representa con la ecuación $Y/(1-Y)$. El modelo logístico es formado por el logaritmo natural (\ln) de esta probabilidad, es decir, $\ln [Y/(1-Y)]$. Al igual que en el cálculo de los intervalos de confianza de 95 % de la razón de probabilidades o de momios (RM), utilizaremos el \ln con el fin de normalizar esta variable dicotómica (el desenlace).

El formato básico de la ecuación de la regresión logística es similar al del modelo de regresión lineal:

$$\ln [Y/(1-Y)] = a + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5$$

Donde Y es el logaritmo natural (\ln) de las probabilidades (*odds*), a es la constante de regresión, e Y es

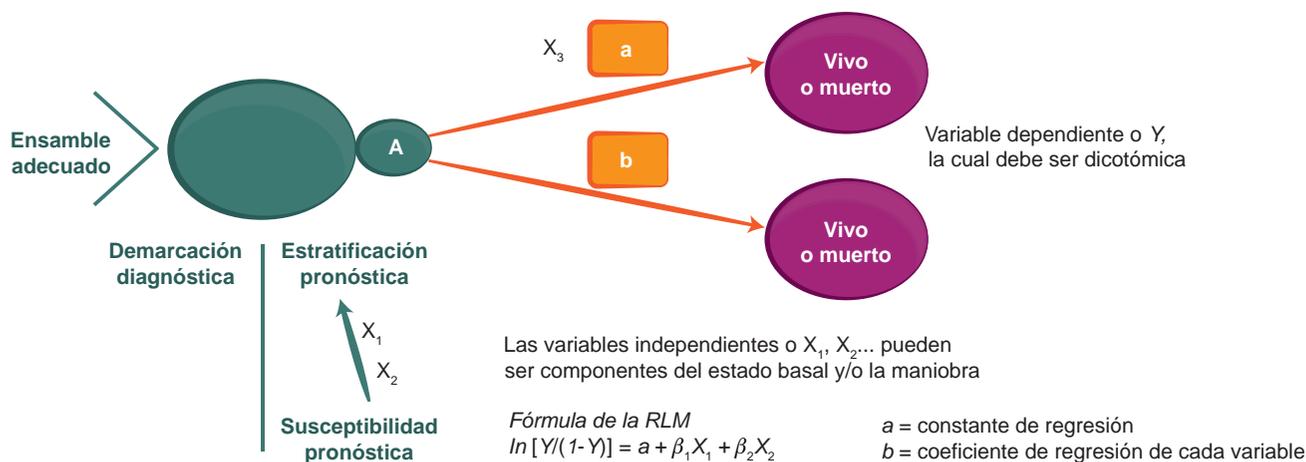


Figura 1 Modelo de regresión logística múltiple dentro del modelo arquitectónico de la investigación

el valor de Y cuando X es igual a cero, lo que significa la probabilidad de la enfermedad cuando no se considera ningún factor de riesgo (y se corresponde con la prevalencia o incidencia de la enfermedad), por ejemplo, la prevalencia de la diabetes mellitus en la población general (16 %). Asimismo, β es el coeficiente de regresión de cada variable y se refiere a la magnitud de cambio en Y por unidad de cambio en X , sin que esto signifique causalidad.

Retomando el ejemplo inicial, supongamos que queremos predecir la posibilidad de que se presente neumonía nosocomial en pacientes de una clínica geriátrica y los catalogamos como sin neumonía (codificados como 0) o con neumonía nosocomial (codificados como 1). Idealmente, en la práctica clínica ubicamos las características basales que al ingreso del paciente pudieran predecir el riesgo de presentar dicha complicación, a fin de tomar medidas preventivas adecuadas. Para esto, se interroga su edad, si tiene otras enfermedades (EPOC, DM, etcétera) o el estado funcional medido con la escala de Karnofsky.

Lo mismo sucede en el análisis de regresión logística múltiple, en el que se recomienda utilizar las variables que en análisis bivariados previos se hubieran mostrado con al menos cierto grado de relación como factores de riesgo ($p < 0.25$ o menor). Debe quedar claro que en el modelo final se pueden incluir todas las variables que han demostrado relación con el desenlace, incluso variables que en el análisis bivariado de nuestros datos no hayan resultado significativas.

También es deseable que cada variable en el modelo cuente con un mínimo de 10 a 20 casos/individuos por cada evento en el menor de los grupos (se consideran dos grupos, el grupo de los sujetos que desarrollan el desenlace y el grupo de los sujetos que no lo desarrollan).

Tomando nuestro ejemplo, si la variable dependiente es neumonía y tuvimos 47 eventos en 121 pacientes, no deberíamos incluir en nuestro modelo de regresión más de cuatro variables ($47/10 = 4.7$). Esta estrategia se utiliza para estabilizar los datos y es llamada comúnmente como *eventos por variable* (*event per variable* o *EPE value*). En caso de incluir más variables, los datos pueden volverse inestables y podemos obtener resultados estadísticamente significativos, sin que en realidad lo sean.

Las preguntas en nuestro ejemplo serían: ¿qué características de los pacientes predicen la aparición de neumonía?, ¿cuál es la probabilidad de que un individuo que ingresa a la clínica geriátrica presente neumonía nosocomial, dada la presencia combinada de dichos factores?

La realización de un modelo de regresión logística múltiple es un proceso complejo, motivo por el que es necesario apoyarnos en programas estadísticos como

SPSS, SAS o Stata, entre otros. En este ejemplo, nos apoyaremos en el programa estadístico SPSS para realizar el análisis.

El primer paso es hacer el análisis bivariado de los datos con el que contrastamos cada una de las posibles variables involucradas contra la variable de desenlace (aparición de neumonía). Las variables dicotómicas (EPOC y diabetes mellitus) se contrastan con χ^2 (capítulo XVII de esta serie), se busca su medida de asociación con el cálculo de la razón de momios (o exponente de β : $\text{Exp}[\beta]$) y el intervalo de confianza (IC 95 %) (capítulo VI de esta serie), mientras que la edad se contrasta con t de Student para grupos independientes (capítulo XV de esta serie).

Una vez que identificamos las variables que se asocian, procederemos a realizar el modelo multivariado (preseleccionar las variables que se van a incluir a través de un análisis bivariado es solo una estrategia, pero bien pueden incluirse todas las variables) e intentar definir cuál de estas variables se relaciona de modo independiente y cuál no.

Como primer paso para realizar la regresión logística múltiple en el programa SPSS, seleccionaremos en la barra de herramientas la opción *Analizar*, posteriormente la pestaña de *Regresión* y finalmente la opción *Logística binaria* (figura 2).

En el cuadro *Dependientes* colocaremos nuestra variable dependiente, en este caso la presencia o no de neumonía. En la sección *Covariables* agregaremos las variables independientes que queremos utilizar para nuestro modelo de predicción. En *Método* asignamos cómo queremos que se lleve a cabo el análisis: *Introducir* (analiza todas las variables incluidas y las deja en el modelo final aun cuando no sean estadísticamente significativas), *Hacia adelante* (automáticamente se irán agregando al modelo una a una las variables significativas estadísticamente) o *Hacia atrás* (del total de variables se irán eliminando automáticamente las que no contribuyan al modelo por falta de significación estadística). En este ejemplo utilizaremos el método *Hacia atrás*.

Posteriormente, tenemos que especificarle al programa las variables que son categóricas, especificar el tipo de contraste (en este caso, simple, en el que cada categoría —ordinal o dicotómica— se contrasta contra la categoría de referencia) y poner de acuerdo con lo observado si la categoría de referencia es la primera o la última (el orden afectará la manera como se representa la RM en la tabla de resultados, ya sea como factor de riesgo o de protección, pero no modifica el resultado del modelo).

Al realizar el análisis, como primer dato se presenta una tabla de clasificación aún sin valores pronosticados y que establece el porcentaje de pacientes que no presentaron neumonía (61.2 %).



Figura 2 Interfaz de inicio para regresión logística en el programa estadístico SPSS

En la figura 3 se muestran las variables de la ecuación. Tenemos las columnas del coeficiente de regresión o *B* (indica el incremento en la variable dependiente con cada unidad de la variable independiente), de la significación estadística y de la RM o $Exp(\beta)$ (indica la probabilidad de desarrollar el evento), con su respectivo intervalo de confianza del 95 %. En este ejemplo tenemos que el logaritmo de posibilidades disminuiría -0.022 con cada año de edad (es decir, a menor edad, mayor probabilidad de desarrollar neumonía durante la hospitalización), aunque al ver que el valor de *p* es > 0.05 y que el intervalo de confianza del $Exp(\beta)$ o la RM atraviesa la unidad, notamos que carece de significación estadística. Por otro lado, tenemos el caso de la EPOC: en los pacientes que tienen esta patología, sus probabilidades aumentan 2.203 ($Exp[B]$), lo que en la clínica se traduce como $RM = 9.05$ ($IC\ 95\ \% = 3.44-23.76$). Por lo tanto, se puede interpretar esta RM como que los que tienen EPOC tienen 8.05 veces más riesgo de tener neumonía en comparación con los que no la

padecen (capítulo VI de esta serie) y que esta asociación es estadísticamente significativa, y se relaciona independientemente de la edad, de la presencia de diabetes mellitus y del Karnofsky.

En este primer paso encontramos que ni la edad, ni el antecedente de tener DM se relacionan con el desarrollo de la neumonía nosocomial y que las variables que se relacionan son tener EPOC y el estado funcional medido con el Karnofsky (figura 3).

En un segundo paso de la regresión logística se pueden eliminar variables con las que no se explique la variable dependiente. Esto se hace al eliminar la variable con mayor valor de *p*, en este caso la presencia de diabetes ($p = 0.611$), por lo que los resultados se pueden ver en la tabla de clasificación y los resultados de variables de la figura 4. Con esto se mejora el porcentaje global de efectividad a 70.2 %. Nosotros recomendamos no eliminar las variables clínicas que tengan lógica biológica (en este ejemplo, la edad), con el fin de ajustar el modelo.

Variables de la ecuación

	B	E.T.	Wald	Sig.	Exp(β)	IC 95% para Exp(β)	
						Inferior	Superior
Paso 1 Edad	-0.022	0.019	1.418	0.234	0.978	0.943	1.014
EPOC(1)	2.203	0.492	20.027	0.000	9.053	3.449	23.761
Diabetes(1)	.301	0.593	0.258	0.611	1.351	0.423	4.318
Karnofsky(1)	1.933	0.677	8.157	0.004	6.912	0.834	26.049
Constante	1.429	1.182	1.461	0.227	4.175		

Figura 3 Resultados del modelo de regresión logística múltiple, en el que la variable dependiente es el desarrollo de neumonía nosocomial

Por ejemplo, si realizamos un tercer paso y se procede a eliminar la variable con p mayor, en este caso la variable edad ($p = 0.195$), se logra un porcentaje global de explicación de 71.1 % (figura 5). Sin embargo, este no es clínicamente mayor al modelo anterior que predecía un 70 %, pero incluía una variable que clínicamente resultaba muy significativa como la edad, y que en estudios previos había demostrado impactar en el desenlace. Si en este estudio no resultó significativa, habrá que ver si tenemos un grupo con una edad muy compacta o si no incluimos grupos de edades similares a los estudios en los que sí resultó significativa.

Tabla de clasificación^a

Observado	Pronosticado		
	Neumonía		% correcto
	sin neumonía	con neumonía	
Paso 2	51	23	68.9
	13	34	72.3
			70.2

^aEl valor de corte es 0.500

Variables en la ecuación

	B	E.T.	Wald	Sig.	Exp(β)	IC 95% para Exp(β)	
						Inferior	Superior
Paso 2 Edad	-0.024	0.018	1.681	0.195	0.976	0.942	1.012
EPOC(1)	2.203	0.492	20.053	0.000	9.055	3.452	23.753
Karnofsky(1)	1.925	0.674	8.148	0.004	6.852	1.828	25.686
Constante	1.624	1.115	2.121	0.145	5.075		

Figura 4 Segundo paso de la regresión logística

Con este modelo podríamos decir que los pacientes con EPOC y Karnofsky de 70 % tienen mayor riesgo de presentar neumonía durante una hospitalización, lo cual puede tener implicaciones en el pronóstico (o en el manejo) al ser considerados como pacientes con mayor riesgo que los que no tienen estas características clínicas.

la variable dependiente, explicada por las variables independientes (ver capítulo VI de esta serie). En este caso, el valor es de 0.213, lo cual significa que el 21 % de la variación de la variable dependiente (neumonía) se explica con las variables independientes incluidas en el modelo (edad, capacidad funcional y EPOC en este ejemplo). El R^2 de Nagelkerke es una versión corregida del R^2 de Cox y Snell.

Resumen del modelo

En los resultados de la prueba de regresión logística se nos presentará el resumen del modelo (figura 6).

El valor de menos dos veces el logaritmo ($-2 \log$) de la verosimilitud, o también llamado desviación, indica hasta qué punto el modelo se ajusta bien a los datos (en él se considera que cuanto más pequeño es el valor, mejor es el ajuste). El R^2 de Cox y Snell es un coeficiente de determinación generalizado que se utiliza para estimar la proporción de varianza de

Tabla de clasificación^a

Observado	Pronosticado		
	Neumonía		Porcentaje correcto
	sin neumonía	con neumonía	
Paso 3	45	29	60.8
	6	41	87.2
			71.1

^aEl valor de corte es 0.500

Variables en la ecuación

	B	E.T.	Wald	Sig.	Exp(β)	IC 95% para Exp(β)	
						Inferior	Superior
Paso 3 EPOC(1)	2.136	0.484	19.462	0.000	8.469	3.278	21.881
Karnofsky(1)	2.076	0.668	9.655	0.002	7.972	2.152	29.532
Constante	239	0.301	0.627	0/428	1.270		

Figura 5 Tercer paso de la regresión logística

Prueba de Hosmer-Lemeshow

En la presentación final de los datos de regresión logística es deseable que figure algún tipo de bondad de ajuste, como la de Hosmer-Lemeshow. Esta prueba analiza, de acuerdo con deciles de riesgo, la presencia del evento contra la frecuencia esperada (probabilidad de neumonía < 10 %, < 20 % y hasta 100 %). Ambas distribuciones, la esperada y la observada, se contrastan con una prueba de χ^2 (figura 7).

Dado lo anterior, podemos inferir que no existe diferencia estadística entre la distribución esperada y la que predice nuestro modelo, por lo que se puede considerar como adecuado para establecer el riesgo de ocurrencia del evento de interés.

Comentarios

El modelo de regresión logística múltiple es una poderosa herramienta para el análisis multivariado y para ponderar una variable dependiente frente a otras con las cuales pudiera interactuar. Es uno de los modelos que más se parece al pensamiento clínico, en el que se conoce que no existe el modelo unicausal, sino que siempre hay múltiples causas y estas tienen distinto peso.

Cuando interpretamos o usamos la regresión logística múltiple debemos tomar en cuenta los siguientes puntos: *a*) la variable dependiente (*Y*) debe ser dicotómica y mutuamente excluyente; *b*) las variables

Paso	-2 log de la verosimilitud	R cuadrado de Cox y Snell	R cuadrado de Nagelkerke
1	132.492	0.214	0.291
2	132.754	0.213	0.288
3	134.476	0.201	0.273

Figura 6 Resumen del modelo de regresión logística

Paso	Chi cuadrado	gl	Sig.
1	3.209	8	0.921
2	3.724	8	0.881
3	0.327	1	0.568

Figura 7 Prueba de Hosmer-Lemeshow

independientes pueden ser de cualquier tipo, ya sean cualitativas o cuantitativas; *c*) los modelos finales deben diseñarse tanto a partir del análisis bivariado como por lógica biológica.

Declaración de conflicto de interés: los autores han completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no ha sido reportado alguno que esté relacionado con este artículo.

^aDepartamento de Medicina Interna, Hospital de Oncología

^bCentro de Adiestramiento en Investigación Clínica (CAIC), Coordinación de Investigación en Salud

^cSubdirección de Investigaciones Clínicas, Instituto Nacional de Psiquiatría "Dr. Ramón de la Fuente Muñiz", Secretaría de Salud

^dDepartamento de Urología, Hospital de Especialidades

^{a,b,d}Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social

Distrito Federal, México

Comunicación con: Ricardo Berea-Baltierra

Correo electrónico: ricberbal@hotmail.com

Referencias

- Portney LG, Watkins MP. Logistic regression. En: Foundations of clinical research applications to practice. Third edition. New Jersey, USA: Pearson & Prentice Hall; 2009. p. 696-700.
- Palacios-Cruz L, Pérez M, Rivas-Ruiz R, Talavera JO. Investigación clínica XVIII. Del juicio clínico al modelo de regresión lineal. Rev Med Inst Mex Seguro Soc. 2013;51(6):656-61.
- Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373-9.
- Dawson B, Trapp RG. Bioestadística médica. México: Manual Moderno; 2005. p. 239-41.
- Talavera JO, Rivas-Ruiz R, Pérez-Rodríguez M. Investigación clínica VI. Relevancia clínica. Rev Med Inst Mex Seguro Soc. 2011;49(6):631-5.
- Feinstein AR. Multivariable analysis: An introduction. New Haven: Yale University Press; 1996.
- Talavera JO. Investigación clínica I. Diseños de investigación. Rev Med Inst Mex Seguro Soc. 2011; 49(1):53-8.
- Rivas-Ruiz R, Castelán-Martínez OD, Pérez M, Talavera JO. Investigación clínica XVII. Prueba chi cuadrada, de lo esperado a lo observado. Rev Med Inst Mex Seguro Soc. 2013;51(5):552-7.
- Rivas-Ruiz R, Pérez-Rodríguez M, Talavera JO. Investigación clínica XV. Del juicio clínico al modelo estadístico. Diferencia de medias. Prueba t de Student. Rev Med Inst Mex Seguro Soc. 2013;51(3): 300-3.



Investigación clínica XXI

Del juicio clínico al análisis de supervivencia

Rodolfo Rivas-Ruiz,^a Marcela Pérez-Rodríguez,^a Lino Palacios,^b Juan O. Talavera^a

...Para aquella fecha habré fallecido, pero la posteridad imparcial no rehusará admitir que este trabajo lo hizo primeramente un inglés...

EDMUNDO HALLEY (1656-1742)

Clinical research XXI. From the clinical judgment to survival analysis

Decision making in health care implies knowledge of the clinical course of the disease. Knowing the course allows us to estimate the likelihood of occurrence of a phenomenon at a given time or its duration. Within the statistical models that allow us to have a summary measure to estimate the time of occurrence of a phenomenon in a given population are the linear regression (the outcome variable is continuous and normally distributed —time to the occurrence of the event—), logistic regression (outcome variable is dichotomous, and it is evaluated at one single interval), and survival curves (outcome event is dichotomous, and it can be evaluated at multiple intervals). The first reference we have of this type of analysis is the work of the astronomer Edmond Halley, an English physicist and mathematician, famous for the calculation of the appearance of the comet orbit, recognized as the first periodic comet (1P/Halley's Comet). Halley also contributed in the area of health to estimate the mortality rate for a Polish population. The survival curve allows us to estimate the probability of an event occurring at different intervals. Also, it leads us to estimate the median survival time of any phenomenon of interest (although the used term is survival, the outcome does not need to be death, it may be the occurrence of any other event).

Keywords Palabras clave

Clinical evolution	Evolución clínica
Survival	Supervivencia
Kaplan-Meier estimate	Estimación de Kaplan-Meier
Life tables	Tablas de vida
Time	Tiempo

La toma de decisiones durante la atención médica implica el conocimiento del curso clínico de la enfermedad. La preocupación por estimar el tiempo que transcurre hasta un evento se ha dado desde mucho tiempo atrás; incluso, las primeras descripciones de tal ejercicio matemático se realizaron en otras áreas del conocimiento, como la física y la astronomía. La primera referencia que se tiene sobre este tipo de análisis es el trabajo publicado en 1693 por el astrónomo, físico y matemático inglés Edmundo Halley —famoso por el cálculo de la órbita y aparición del cometa reconocido como el primer cometa periódico (el 1P/el cometa Halley)—, quien contribuyó en el área de la salud con la estimación de la tasa de mortalidad para una población polaca.

El conocimiento de la evolución clínica de la enfermedad permite estimar la posibilidad de ocurrencia de un fenómeno en un tiempo determinado o su tiempo de duración, por ejemplo, ¿cuánto tiempo se requiere para que se eleven los anticuerpos después de aplicar una vacuna y la persona vacunada quede protegida contra la enfermedad?, ¿cuánto tiempo dura un cuadro de influenza? o ¿cuánto tiempo sobrevive un paciente con cáncer (por supuesto dependiendo del tipo y estadio)?

Entre los modelos estadísticos que permiten tener una medida de resumen para estimar el tiempo de ocurrencia de un fenómeno en una población determinada se encuentran la regresión lineal (variable dependiente continua y con distribución normal [tiempo a la ocurrencia del fenómeno]),¹ la regresión logística (variable dependiente dicotómica, en un solo intervalo de tiempo fijo)² y las curvas de supervivencia (variable dependiente dicotómica en múltiples intervalos de tiempo).

La regresión lineal aplica solo si todos los pacientes llegan al evento final; se requiere este dato puesto que maneja como desenlace una variable continua (por ejemplo, tiempo o muerte) y no puede manejar un desenlace que solo una porción ha sufrido y la otra no.

En el modelo de regresión logística, el investigador determina —con base en el conocimiento previo— el tiempo en el que se estimará la probabilidad de que ocurra un fenómeno. Sin embargo, las curvas de supervivencia permiten calcular la probabilidad de que ocurra un fenómeno en distintos intervalos, de forma

^aCentro de Adiestramiento en Investigación Clínica (CAIC), Coordinación de Investigación en Salud, Instituto Mexicano del Seguro Social

^bSubdirección de Investigaciones Clínicas, Instituto Nacional de Psiquiatría "Dr. Ramón de la Fuente Muñiz", Secretaría de Salud

Distrito Federal, México

Comunicación con: Rodolfo Rivas-Ruiz
Correo electrónico: rivasrodolfo@gmail.com

Resumen

La toma de decisiones durante la atención médica implica el conocimiento de la evolución clínica de la enfermedad. A su vez, el conocimiento de la evolución clínica permite estimar la posibilidad de ocurrencia de un fenómeno en un tiempo determinado o su tiempo de duración. Entre los modelos estadísticos con los que es posible obtener una medida de resumen para estimar el tiempo de ocurrencia de un fenómeno en una población determinada se encuentran la regresión lineal (variable dependiente continua y con distribución normal [tiempo a la ocurrencia del fenómeno]), la regresión logística (variable dependiente dicotómica, en un solo intervalo) y las curvas de supervivencia (variable dependiente dicotómica, en múltiples intervalos). La primera referen-

cia que se tiene sobre este tipo de análisis es el trabajo publicado en 1693 por el astrónomo, físico y matemático inglés Edmundo Halley —famoso por el cálculo de la órbita y aparición del cometa reconocido como el primer cometa periódico (el 1P/cometa Halley)— quien contribuyó en el área de la salud con la estimación de la tasa de mortalidad para una población polaca. Las curvas de supervivencia permiten calcular la probabilidad de que ocurra un fenómeno a distintos intervalos y, de igual forma, permiten estimar la mediana de supervivencia de cualquier fenómeno de interés (aunque se utiliza el término supervivencia, el desenlace no necesariamente tiene que ser muerte, sino la ocurrencia de cualquier otro fenómeno).

similar a la regresión logística, aunque esta última a un solo intervalo. Las curvas de supervivencia también permiten estimar la mediana de supervivencia de cualquier fenómeno de interés (aunque se utiliza el término supervivencia, el desenlace no necesariamente tiene que ser muerte, puede ser la ocurrencia de cualquier otro fenómeno), de manera análoga al modelo de regresión lineal, pero en este caso el resultado es la mediana de supervivencia, no la media.

A diferencia de la regresión logística y la regresión lineal, las curvas de supervivencia hacen posible el estudio de sujetos por periodos parciales y no es indispensable seguirlos hasta la ocurrencia del fenómeno o la terminación del tiempo de seguimiento.

Un evento de interés puede referirse a la ocurrencia de una complicación en el diabético, la recaída de una neoplasia, la curación o la muerte.³ La característica

principal de este evento es su naturaleza dicotómica (presente o ausente). Cuando el sujeto no llega al evento de interés y se cuenta solo con un seguimiento parcial del mismo, dentro del análisis se le denomina “dato censurado” (figura 1). En general existen tres razones para “censurar” un dato:³

- Si un sujeto no presenta el evento antes del término del estudio.
- Si no es posible el seguimiento del sujeto durante el transcurso del estudio (pérdida de seguimiento).
- Si una persona decide renunciar al estudio (o muere, siempre y cuando la muerte no sea el evento de interés).

Con la finalidad de lograr una mayor precisión para determinar el tiempo de ocurrencia del desenlace de

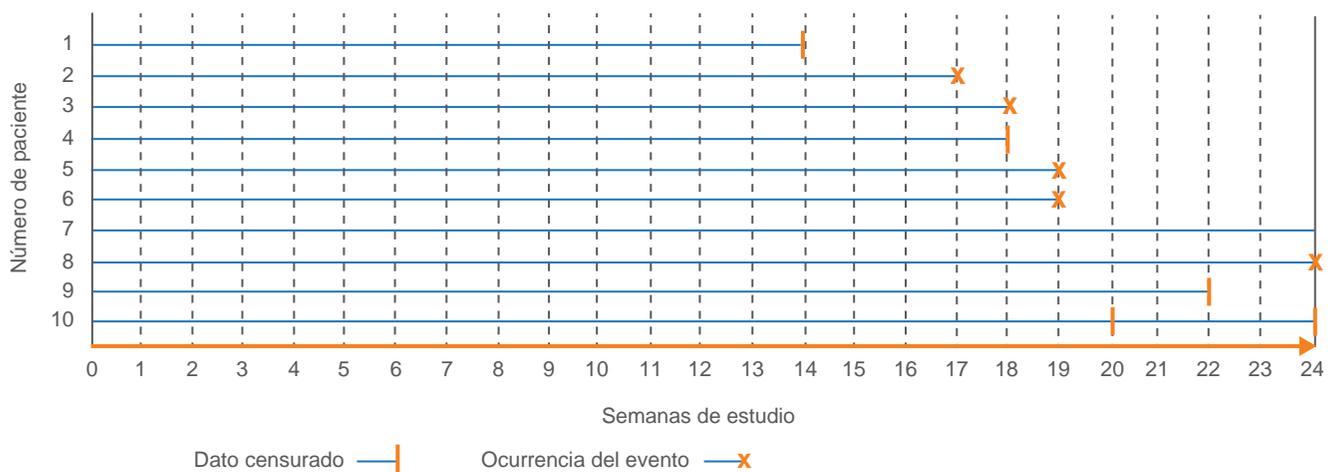


Figura 1 Representación gráfica del seguimiento de pacientes para el análisis de supervivencia. Cada línea horizontal con un número a la izquierda (1 al 10) representa un paciente y su seguimiento a lo largo de los 24 meses que duró un estudio. Los participantes 1, 4, 9 y 10 se consideran datos “censurados” ya que tienen un seguimiento incompleto. Los pacientes 2, 3, 5, 6 y 8 presentaron el evento de interés a lo largo del estudio y el participante 7 terminó el periodo de seguimiento sin haber sido censurado y sin haber presentado el evento de interés.

interés, en el análisis de supervivencia se programan visitas periódicas.

El seguimiento de los sujetos inicia en la fecha en que cada uno se incorporó al estudio, y dicho seguimiento puede o no llegar a la fecha de cierre del mismo. La fecha de inicio representa el estado basal de cada paciente y, en consecuencia, es la misma para todos ellos, no importa que se incluyan en fechas-calendario distintas (por ejemplo, unos pacientes en marzo y otros en abril). Las fechas de término también pueden ser diferentes, dependiendo de si se trata de un sujeto censurado o del momento en que presente el desenlace, o si se termina el tiempo de seguimiento sin que se haya presentado el desenlace o el evento de interés.

Los requisitos para disponer de datos adecuados para un análisis de supervivencia son los siguientes:

- Definir apropiadamente el origen o inicio del seguimiento.
- Definir apropiadamente la escala del tiempo.
- Definir apropiadamente el evento.

El origen o inicio del seguimiento tiene que establecerse antes de la ocurrencia del desenlace, de preferencia cuando apenas ha iniciado la exposición a la maniobra o antes de la misma (maniobra observacional en el caso de una cohorte, o experimental en el caso de un ensayo clínico).

Las escalas de medición del tiempo de seguimiento pueden ser de intervalos muy variables: de minutos

cuando se trata de eventos agudos o de años cuando se analizan patologías de larga duración. Un ejemplo: cuando ocurre un infarto cerebral resulta relevante el tiempo que transcurre para el inicio del tratamiento fibrinolítico, porque es vital que no rebase los 180 minutos y mientras menos tiempo transcurra, el pronóstico del paciente es mejor.

Finalmente está la definición del evento de interés (desenlace). Si bien el análisis estadístico recibe el nombre de curvas de supervivencia, no siempre se refiere a la pérdida de la vida, puede aludir a la pérdida de la salud, a la aparición de un evento adverso, a enfermedad o complicación de la misma. Además, la representación gráfica de los resultados puede ser en forma descendente, como tradicionalmente se hace para mostrar la pérdida de la vida, partir del total de los sujetos vivo (100 %) y en cada periodo indicar la proporción de ellos que aún continúa con vida; o bien, en forma ascendente si el objetivo es mostrar la aparición de una enfermedad o sus complicaciones o algún evento adverso, en este caso se parte de cero y en cada periodo se muestra la proporción de sujetos que hasta ese momento han presentado el evento.

El análisis inicial se realiza mediante la creación de tablas de supervivencia, las cuales constan de un tiempo de seguimiento (tiempo de exposición a la maniobra, ya sea observacional o experimental) y un desenlace, y en las que se asume que todos los pacientes inician en condiciones homogéneas en el estado basal.

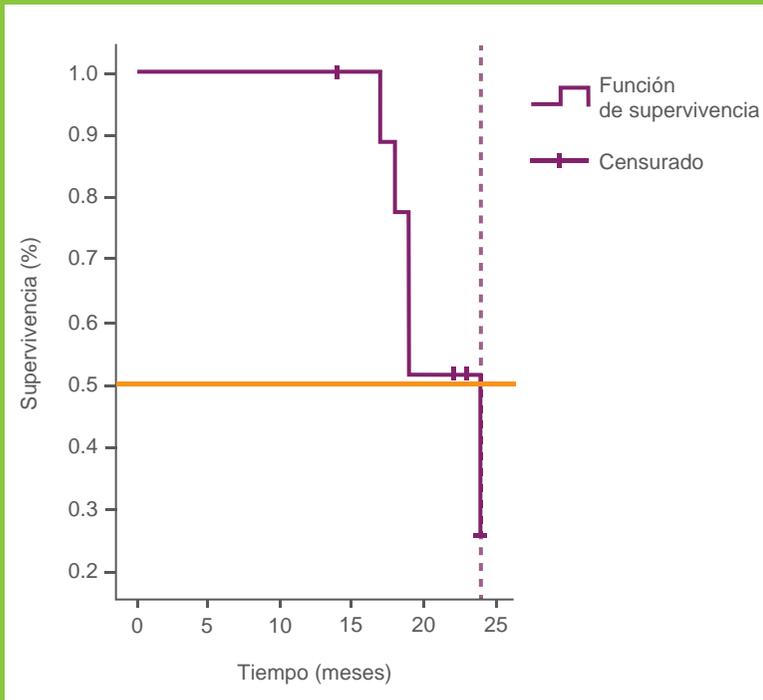
Ejemplo de análisis de supervivencia

Estudio de 10 pacientes con cáncer de mama, las cuales son seguidas durante un período de dos años (24 meses)

Paciente	Tiempo de entrada al estudio	Tiempo de presentación del desenlace	Estado de salida
1	0	14	Censurado (vivo) pérdida de seguimiento
2	0	17	Muerte
3	0	18	Muerte
4	0	18	Censurado (murió por otras causas)
5	0	18	Muerte
6	0	19	Muerte
7	0	24	Vivo
8	0	24	Muerte
9	1	23	Censurado
10	2	22	Censurado

La paciente 1 acudió a sus consultas de seguimiento solo durante 14 meses. No es posible determinar exactamente qué sucedió con ella, solo que estaba viva la última vez que asistió a consulta.

En las pacientes que murieron se registra la fecha de defunción y se calcula el tiempo que estuvieron vivas durante el estudio. En la paciente número 4, que falleció por causas ajenas al cáncer (por un accidente automovilístico), se registra la fecha de defunción, pero no como evento acumulado sino como caso "censurado". Es necesario valorar si una muerte, aun sin ser ocasionada directamente por el cáncer, clínicamente está relacionada, para de esta forma considerarla como evento. Hubo una paciente que al terminar el estudio continuaba viva (paciente 7).



Las mujeres que fallecieron o que presentaron eventos provocan una línea hacia abajo que muestra la proporción de supervivientes. Por su parte, los casos censurados se muestran con una cruz (cuatro pacientes que no murieron, pero no terminaron sus 24 meses de seguimiento) y una mujer que continuaba viva al término del tiempo de seguimiento.

Intervalo en meses	Expuestos	Censurados	Eventos	Probabilidad de supervivencia del intervalo	Probabilidad de supervivencia acumulada
14	10	1	0	10/10 (100 %)	100.0 %
17	9	0	1	8/9 (89 %)	89.0 %
18	8	1	1	7/8 (88 %)	78.3 %
19	6	0	2	4/6 (66 %)	51.7 %
20	4	1	0	4/4 (100 %)	51.7 %
22	3	1	0	3/3 (100 %)	51.7 %
24	2	0	0	1/2 (50 %)	25.9 %

Se muestran los datos de ocurrencia del evento y la probabilidad de supervivencia en cada periodo y la probabilidad acumulada de supervivencia conforme avanza el seguimiento: en la semana 14 de seguimiento, la probabilidad de supervivencia fue de 100%; al mes 17 fue de 89%, porque hubo un evento en las nueve mujeres que estaban expuestas (en seguimiento). En el mes 18 quedaban ocho mujeres personas expuestas y se presentó un evento, por lo que la probabilidad de supervivencia fue de 88%, sin embargo, la probabilidad acumulada de supervivencia fue de 78.3%.

La probabilidad acumulada se obtiene multiplicando la probabilidad de supervivencia del periodo actual x la probabilidad de supervivencia acumulada hasta ese momento, por ejemplo:

Probabilidad acumulada al mes 24

51.7 % (probabilidad acumulada hasta la semana 22) × 50 % (probabilidad de supervivencia del periodo en cuestión) = 25.9 %.



Cuadro I Media del tiempo de supervivencia

		Media ^a	
Estimación	Error típico	Intervalo de confianza de 95 %	
		Límite inferior	Límite superior
21.778	1.320	19.191	24.365

^aLa estimación se limita al mayor tiempo de supervivencia si se ha censurado

Para determinar el tiempo de supervivencia se pueden utilizar dos métodos: el método actuarial o la técnica publicada por Kaplan y Meier en 1958. Actualmente, con la proliferación y mejora de las computadoras y los programas estadísticos, el método de Kaplan-Meier es el más usado, por ser más gráfico y descriptivo que el método actuarial del cual se obtienen tasas de incidencia o densidad de incidencia.

Cuadro II Mediana del tiempo de supervivencia

		Mediana ^a	
Estimación	Error típico	Intervalo de confianza de 95 %	
		Límite inferior	Límite superior
21.778	1.320	19.191	24.365

^aLa estimación se limita al mayor tiempo de supervivencia si se ha "censurado"

Análisis de Kaplan-Meier

En 1958, Edward Kaplan y Paul Meier publicaron un artículo llamado "Non parametric incomplete estimation from observations",⁴ en el que describen la función matemática para realizar el modelo de curvas de supervivencia. Debido a la complejidad del análisis, fue hasta el *boom* de la computación cuando se dio el mayor auge de esta prueba estadística no paramétrica.

El estimador de Kaplan-Meier, también llamado del producto límite, es un método estadístico no paramétrico en el que se toman en cuenta cada uno de los tiempos que aporta cada paciente. Se utiliza cuando se conocen los tiempos individuales de los "censurados" y "no censurados", de manera que se calcula la supervivencia cada vez que un paciente muere o alcanza el tiempo de seguimiento. Permite calcular la mediana de supervivencia, lo que significa, cuánto tiempo tarda 50 % de la población para llegar al evento. Además, permite calcular los valores del intervalo de confianza de 95 % (IC 95 %) de esta estimación:

$$IC\ 95\ \% = \text{error estándar o típico} \times \pm 1.96 (Z\ \alpha\ \text{deseado})$$

Si se desea una confianza de 95 %, el valor crítico de $Z\ \alpha$ sería 1.96. Al multiplicar el error estándar $\times 1.96$ se obtiene el IC 95 % superior y al multiplicar el error estándar $\times -1.96$ se obtiene el IC 95 % inferior.

Cuando la población estudiada no alcanza 50 % del resultado esperado, se determina la media de supervivencia, siempre y cuando esta tenga distribución normal (cuadro I). Sin embargo, cuando el seguimiento es suficiente para conocer cuál es el resultado (desenlace en 50 %), se calcula la mediana de supervivencia (cuadro II).

Es preferible usar la mediana de supervivencia (en lugar de la media), la cual puede ser interpretada como el tiempo que tarda 50 % de la población en llegar al evento. Ambos valores pueden coincidir. Cuando se realiza una comparación, el método de Kaplan-Meier calcula inicialmente los tiempos de supervivencia, tanto la mediana como la media (cuadro III).

La validez de este método se fundamenta en dos suposiciones:

- Las personas que se retiran del estudio tienen un destino parecido a las que quedan.
- El tiempo-calendario durante el cual una persona entra en el estudio no tiene efecto independiente en la aparición del evento.

Cuadro III Medias y medianas del tiempo de supervivencia obtenidas con el método de Kaplan-Meier

Grupo	Media ^a				Mediana			
	Estimación	Error típico	IC 95 %		Estimación	Error típico	IC 95 %	
			Límite inferior	Límite superior			Límite inferior	Límite superior
Tratamiento A	21.778	0.880	20.053	23.502	25.000	1.941	21.196	28.804
Tratamiento B	11.278	0.917	9.481	13.074	9.000	1.797	5.479	12.521
Global	18.194	1.176	15.889	20.498	19.000	.527	17.967	20.033

^aLa estimación se limita al mayor tiempo de supervivencia si se ha "censurado"

IC 95 % = intervalo de confianza de 95 %

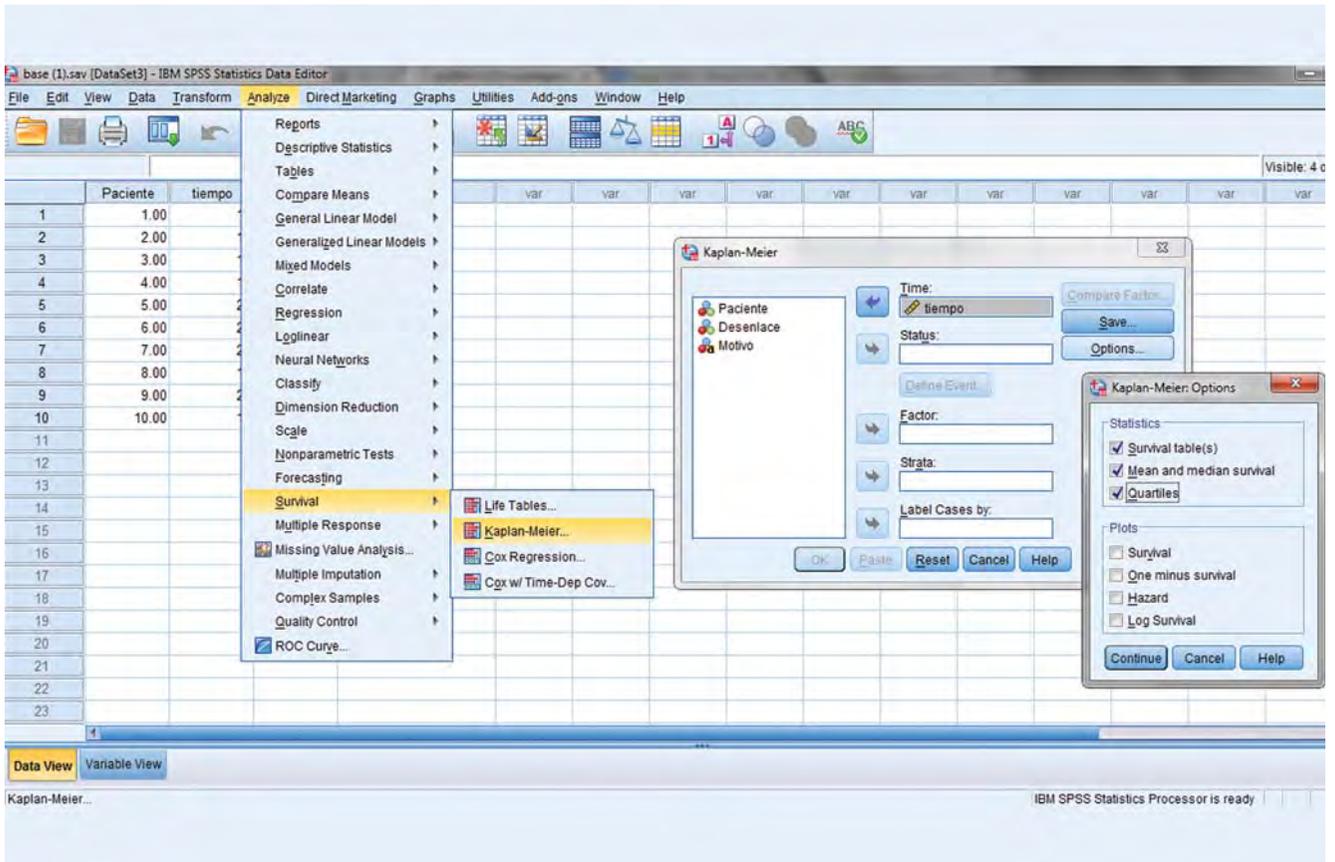


Figura 2 Pasos para la realización del análisis de Kaplan-Meier en el programa estadístico SPSS

Los pasos para realizar este análisis en el programa estadístico SPSS se muestran en la figura 2. Se comienza seleccionando la opción “Analizar”, luego se selecciona la función “Supervivencia” y finalmente se selecciona la opción “Kaplan-Meier”. Después se selecciona la variable que contenga el tiempo de seguimiento y se agrega a la casilla de tiempo, posteriormente se selecciona la variable de desenlace y se agrega a la casilla de estatus, definiendo el evento (ya sea como 0 o 1; generalmente la presencia del evento se codifica con 1) en la opción que se encuentra debajo de la casilla.

Análisis bivariado

Para comparar el efecto de una intervención en los estudios de supervivencia, como la administración de un medicamento, y conocer si existen diferencias y si estas son mayores a las que se pudieran presentar por el efecto del azar, se utiliza el método de la prueba de logaritmo de rango (*log-rank test*). Esta prueba se basa en el modelo de χ^2 (véase la parte XVI de esta serie⁵), en la cual se comparan los eventos (muerte) en cada grupo. El resultado de esta prueba es un valor de *p* (probabilidad de que las diferencias se deban al azar),

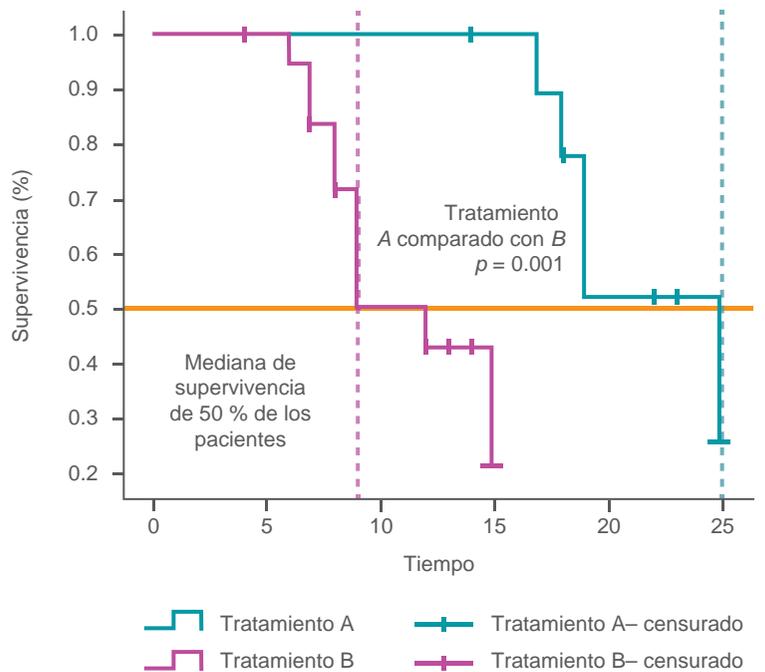


Figura 3 Contraste de las curvas de supervivencia con la prueba de log rank. En esta figura se observan caídas pronunciadas porque la muestra comprende solo a 10 personas, sin embargo, cuando se analizan muestras mayores en lugar de escalones se observa una pendiente

Cuadro IV Análisis de las diferencias globales de los grupos de tratamiento usando la prueba del rango de logaritmo

	χ^2	gl	Sig.
Log rank (Mantel-Cox)	19.611	1	0.000

Prueba de igualdad de distribuciones de supervivencia para diferentes niveles de grupo

además de la mediana o la media de supervivencia. Este análisis también es susceptible de ser graficado y las diferencias entre grupos se muestran al contrastar el efecto en 50 % de la población, es decir, medianas de supervivencia (figura 3).

Además, se puede calcular el valor de *p* a partir de la prueba de la χ^2 con el método de *log-rank*, tomando a los grupos como los grados de libertad (gl) con la siguiente fórmula, donde la *n* son los grupos de tratamiento (cuadro IV):⁶

$$gl = n - 1$$

Los pasos para realizar este análisis en el programa estadístico SPSS se muestran en la figura 4. Al igual que en el análisis de Kaplan-Meier, se comienza selec-

cionando la opción “Analizar”, luego se elige “Supervivencia” y la opción “Kaplan-Meier”. Después se selecciona la variable que contenga el tiempo de seguimiento y la variable de desenlace (evento) y se agrega la variable grupo de tratamiento (en este caso A o B). Para comparar estos grupos se presiona “Comparar factor” y se selecciona “Log rank”. En “Opciones” se pueden solicitar las tablas de supervivencia y los gráficos, estos últimos con opciones como las que presentamos aquí de supervivencia.

Conclusiones

Las curvas de supervivencia permiten estimar la probabilidad de que ocurra un fenómeno en distintos intervalos. La posibilidad de utilizar la información de sujetos que se pierden durante el seguimiento, las hace un modelo extraordinario en el manejo de datos incompletos, sin embargo, siempre habrá que cuidar que el número de sujetos “censurados” sea el menor posible, de otra manera la estimación se realizará a partir de los datos de unos cuantos sujetos como si se tratara de la totalidad del grupo.

Declaración de conflicto de interés: los autores han completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no fue reportado alguno en relación con este artículo.

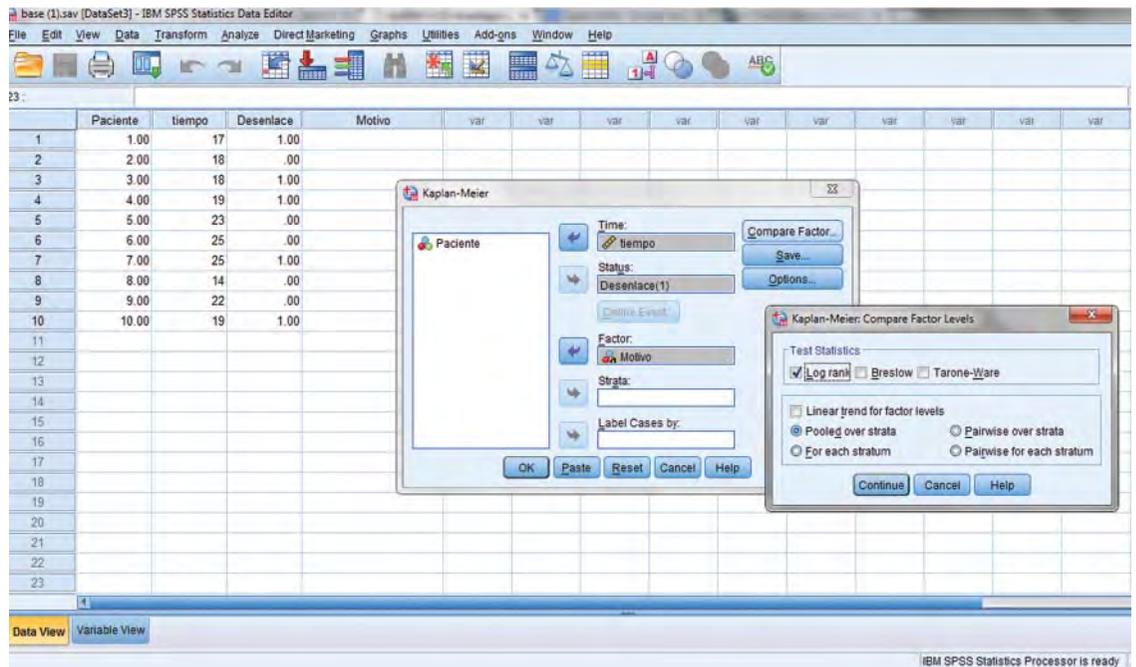


Figura 4 Pasos para la realización de la prueba de *log-rank* en el programa SPSS

Referencias

1. Palacios-Cruz L, Pérez-Rodríguez M, Rivas-Ruiz R, Talavera JO. Investigación clínica XVIII. Del juicio clínico al modelo de regresión lineal. *Rev Med Inst Mex Seguro Soc.* 2013;51(6):656-61.
2. Pérez-Rodríguez M, Palacios-Cruz L, Moreno J, Rivas-Ruiz R, Talavera JO. Investigación clínica XIX. Del juicio clínico al análisis de covarianza. *Rev Med Inst Mex Seguro Soc.* 2014;52(1):70-5.
3. Kleinbaum D. *Survival analysis: A self-learning text.* Second edition. Berlin: Springer Science Business Media; 2005.
4. Kaplan EL, Meier P. Nonparametric incomplete estimation from observations. *J Am Statist Assoc.* 1958; 53(282):457-481.
5. Rivas-Ruiz R, Moreno-Palacios J, Talavera JO. Investigación clínica XVI. Diferencias de medianas con la U de Mann-Whitney. *Rev Med Inst Mex Seguro Soc.* 2013;51(4):414-9
6. Feinstein AR. *Principles of medical statistics.* New York, NY: Chapman and Hall/CRC; 2002.

Lecturas recomendadas

- Abd-ElHafeez S, Torino C, D'Arrigo G, Bolignano D, Provenzano F, Mattace-Raso F, Zoccali C, Tripepi G. An overview on standard statistical methods for assessing exposure-outcome link in survival analysis (Part II): The Kaplan-Meier analysis and the Cox regression method. *Aging Clin Exp Res.* 2012;24(3):203-6.
- Benítez-Parejo N, Rodríguez-del Águila MM, Pérez-Vicente S. *Survival analysis and Cox regression.* *Allergol Immunopathol (Madr).* 2011;39(6):362-73.
- Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). *BMJ.* 1998;317(7172):1572.
- Efron B. Logistic regression, survival analysis, and the Kaplan-Meier curve. *J Am Stat Assoc.* 1988; 83(402): 414-25.
- Satagopan JM, Ben-Porat L, Berwick M, Robson M, Kutler D, Auerbach AD. A note on competing risks in survival data analysis. *Br J Cancer.* 2004;91(7):1229-35.
- Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med.* 2005;24(20):3089-110.



Investigación clínica XXII

Del juicio clínico al modelo de riesgos proporcionales de Cox

Marcela Pérez-Rodríguez,^a Rodolfo Rivas-Ruiz,^a Lino Palacios-Cruz,^b Juan O. Talavera^a

Primero calcula, después, piensa...

SIR DAVID ROXBEE COX

Clinical research XXII. From clinical judgment to Cox proportional hazards model

Survival analyses are commonly used to determine the time of an event (for example, death). However, they can be used also for other clinical outcomes on the condition that these are dichotomous, for example healing time. These analyses only consider the relationship of one variable. However, Cox proportional hazards model is a multivariate analysis of the survival analysis, in which other potentially confounding covariates of the effect of the main maneuver studied, such as age, gender or disease stage, are taken into account. This analysis can include both quantitative and qualitative variables in the model. The measure of association used is called hazard ratio (HR) or relative risk ratio, which is not the same as the relative risk or odds ratio (OR). The difference is that the HR refers to the possibility that one of the groups develops the event before it is compared with the other group. The proportional hazards multivariate model of Cox is the most widely used in medicine when the phenomenon is studied in two dimensions: time and event.

Keywords Palabras clave

Survival	Supervivencia
Multivariate models	Modelos multivariados
Hazard ratio	Hazard ratio

Desde el artículo previo (capítulo XXI de esta serie), cuando hablábamos de curvas de supervivencia, mencionamos que en gran parte del ejercicio clínico relacionado con la atención médica la toma de decisiones implica el conocimiento del curso clínico de la enfermedad, es decir, la estimación del tiempo transcurrido hasta que un evento suceda.

Uno de los ejemplos más comunes es cuando un clínico desea estimar cuánto tiempo podría vivir un paciente con cáncer de pulmón con o sin tratamiento. Para realizar esta estimación se usa el método de Kaplan-Meier. Sin embargo, este método no permite plantear un modelo que tome en cuenta las distintas condiciones del paciente que pudieran modificar el desenlace, como la edad, la etapa clínica, etc. Además asume que los grupos son homogéneos con respecto a todas las variables; por ejemplo, asume que todos los sujetos tienen la misma edad, lo cual no es necesariamente el caso de los pacientes de los cuales queremos emitir un pronóstico confiable que nos permita realizar un juicio clínico adecuado.

A partir de este punto tenemos claro que, y lo hemos señalado previamente al tratar el tema del análisis de regresión logística múltiple (capítulo XX de esta serie), para que un modelo multivariado sea considerado completo debe ponderar el efecto de las distintas variables previamente consideradas, para la predicción de una o más variables dependientes, cuyo desenlace es dicotómico. Para este caso, el modelo de riesgos proporcionales de Cox (MRPC) constituye un modelo multivariado que puede ponderar el efecto de una serie de variables cualitativas o cuantitativas sobre un desenlace dicotómico a través del tiempo.

El MRPC fue creado y publicado en 1972, en la revista *Journal of the Royal Statistical Society* por el estadístico inglés Sir David R. Cox, como una alternativa al método de Kaplan-Meier en la que se incorporan las características del análisis de regresión en las tablas de supervivencia.

La función de riesgo para el MRPC puede expresarse mediante la siguiente ecuación:

$$\lambda(t; Z_i(t)) = \lambda_0(t) e^{\beta Z_i(t)} \text{ ó } \lambda_i(t) = \lambda_0(t) e^{x_i(t) \beta}$$

^aCentro de Adiestramiento en Investigación Clínica (CAIC), Coordinación de Investigación en Salud, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social

^bSubdirección de Investigaciones Clínicas, Instituto Nacional de Psiquiatría "Dr. Ramón de la Fuente Muñiz", Secretaría de Salud

Distrito Federal, México

Comunicación con: Marcela Pérez-Rodríguez
Correo electrónico: marxelapr@gmail.com

Resumen

Los análisis de supervivencia son usados comúnmente para establecer el tiempo de ocurrencia de un evento (por ejemplo, muerte). Sin embargo, pueden ser utilizados para otros desenlaces clínicos siempre y cuando estos sean dicotómicos, como tiempo de curación, tiempo de recaída, tiempo para que una enfermedad inicie, etc. Los análisis de Kaplan-Meier (K-M) solo consideran la relación de una variable a través del tiempo, mientras que los riesgos proporcionales de Cox son el modelo multivariado de este método, el cual toma en cuenta otras covariables posiblemente confusoras del efecto de la maniobra principal

estudiada, como la edad, el sexo o el estadio de la enfermedad. Este análisis puede incluir en su modelo variables dependientes cuantitativas y cualitativas. La medida de asociación que se usa se llama hazard ratio (HR) o razón de riesgos, la cual no es lo mismo que el riesgo relativo o la razón de momios (RM). La diferencia es que el HR se refiere a la posibilidad de que uno de los grupos llegue antes a un evento al compararlo con otro. El modelo de riesgos proporcionales de Cox es el modelo multivariado más usado en la medicina cuando se estudia el fenómeno en dos dimensiones: tiempo y evento.

Donde $\lambda_0(t)$ es una función de riesgo no negativa y sin especificar, común a todos los sujetos del estudio, llamada *función de riesgo base*, y β es el vector de coeficientes del modelo. Es decir, el MRPC se utiliza fundamentalmente para comparar grupos y calcular cocientes de riesgo.

$$\text{Hazard} = \lambda = \frac{\text{Sucesos ocurridos en el instante } t}{\text{Sucesos en riesgo en el instante } t} = \frac{d}{n}$$

Dicho de una manera sencilla, la principal innovación del modelo es ofrecer un método de análisis de supervivencia sin necesidad de especificar o estimar una función de riesgo basal, mediante el desarrollo del concepto de *probabilidad marginal*, una función de probabilidad que solo depende de los coeficientes del modelo. A diferencia de la regresión logística, en este modelo se empieza con el 100 % de los sujetos en una condición específica; por ejemplo, vivo cuando se busca muerte, o sin evento adverso alguno cuando se inicia una terapia.

Como sabemos, cuando alguien muere no podemos atribuirlo a una sola causa, sino a múltiples factores como los que se encuentran al inicio de la enfermedad (o estado basal), los cuales pudieran corresponder a la demarcación pronóstica. Para ejemplificar lo que estamos comentando, pensemos en un caso hipotético en el cual estamos estudiando pacientes con cáncer de pulmón, y en los cuales queremos determinar qué variables se relacionan con el tiempo en que van a morir (variable dependiente o desenlace), tomando en cuenta por un lado que en el estado basal encontraríamos características como edad y sexo de los pacientes, estadio de la enfermedad y tabaquismo. En este caso, para poder estimar un desenlace con mayor certeza, podríamos tomar en cuenta estas variables para determinar el probable tiempo de supervivencia de un paciente, dependiendo de la presencia o ausencia de dichas características.

Al usar este modelo, el resultado se expresa por medio de una estadística llamada hazard ratio (HR), que es una división de riesgo (h) entre grupos:

$$\text{HR} = \frac{\text{Sucesos ocurridos en el instante } t \text{ en el grupo 1} / \text{Sujeto en riesgo en el instante } t \text{ en el grupo 1}}{\text{Sucesos ocurridos en el instante } t \text{ en el grupo 2} / \text{Sujeto en riesgo en el instante } t \text{ en el grupo 2}} = \frac{h1}{h2}$$

El HR, al igual que la razón de momios y el riesgo relativo, es una medida de riesgo; sin embargo, el HR se refiere a la probabilidad que tiene que presentar el desenlace en un periodo de tiempo en comparación con el grupo de referencia o control, es decir, es una probabilidad de tener el desenlace en un periodo de tiempo determinado. En los análisis de supervivencia, la mejor medida de tiempo que se debe considerar es la mediana de tiempo, en la que el 50 % de los participantes habrán desarrollado el desenlace.

Existen tres objetivos estadísticos que típicamente son considerados en el MRPC:

1. La prueba de significación de la maniobra que será la variable de agrupación, por ejemplo, el grupo de tratamiento, edad, estadio.
2. La obtención del estimador del efecto de la maniobra o variable de agrupación, que en este caso es el HR.
3. La obtención del intervalo de confianza para el HR.

Para probar la significación del efecto de la variable de agrupación o de tratamiento se utiliza la prueba de Wald, que proporciona un valor de p . Los supuestos que deben cumplirse en el modelo de Cox son los siguientes:

- Los datos censurados no están relacionados con la probabilidad de que ocurra el evento; es decir, en los sujetos cuyo desenlace desconocemos (sea cual sea la razón por la que se censuraron sus datos), se

presume que la falta de datos no se relaciona con el evento del desenlace.

- Las curvas de supervivencia para cada uno de los estratos (tratamientos u otra variable de agrupación) deben tener funciones de riesgo que sean proporcionales en el tiempo; esto debido a que es un modelo lineal y lo que significa es que cada cambio en el desenlace es proporcional a cada cambio en la variable de riesgo.

Una vez explicado lo anterior, regresemos al ejemplo del estudio en el cual pretendemos predecir el tiempo de muerte para cáncer de pulmón. Supongamos que tenemos una cohorte de 130 pacientes (hom-

bres y mujeres) con cáncer de pulmón en diferentes estadios de la enfermedad y queremos saber cuál es el riesgo de mortalidad en cada uno de esos estadios con base en un ajuste de diferentes datos clínicos que probablemente influyen en el desenlace (muerte), como la edad, el sexo y el tabaquismo.

La figura 1 muestra los pasos para realizar el análisis de riesgos proporcionales de Cox en el programa SPSS:

- El primer paso consiste en seleccionar del menú principal la opción Analizar.
- Sespués seleccionar Supervivencia y posteriormente Regresión de Cox.

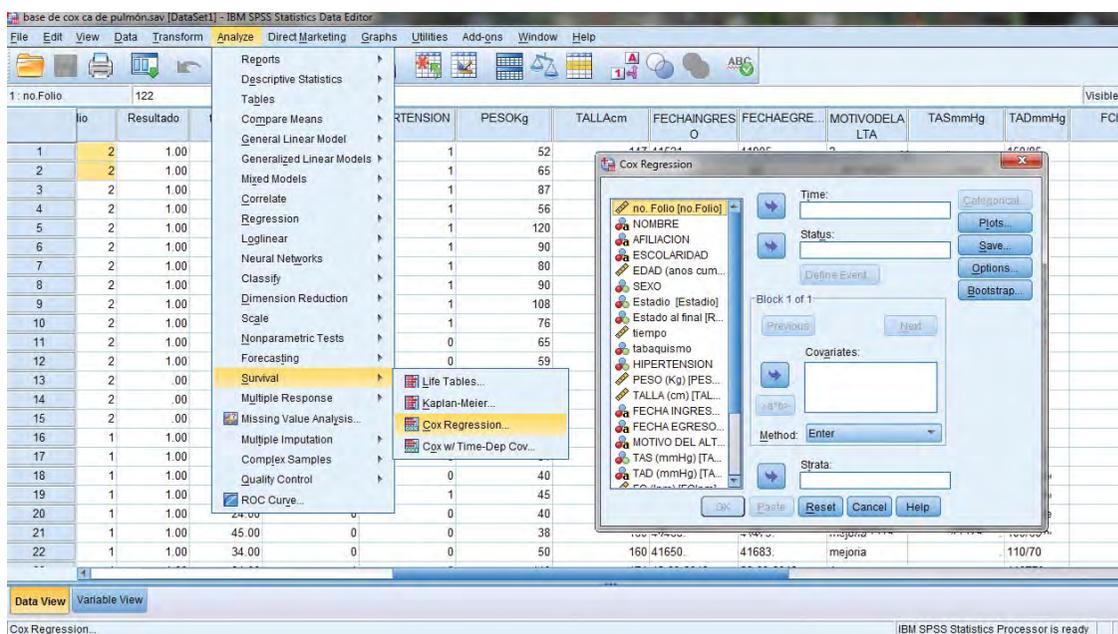


Figura 1 Pasos en el programa estadístico SPSS para realizar el análisis de riesgos proporcionales de Cox

		Frecuencia		
Sexo	1 = Mujer	62	1	
	2 = Hombre	68	0	
Estadio	0 = Etapa I	65	0	0
	1 = Etapa II	50	1	0
	2 = Etapa III	15	0	1
Tabaquismo	0 = ausente	77	0	
	1= presente	53	1	

- Variable de categoría: Sexo
- Codificación de parámetros de indicador
- Variabilidad de categoría: estadio (Estadio)
- Variable de categoría: tabaquismo

Figura 2 Codificación de las variables categóricas^{a,c,d}

En la figura 2 se presentan la frecuencia de las variables categóricas y la forma como se codificaron dichas variables para el análisis.

Una vez que está abierto el cuadro de diálogo (figura 3), el siguiente paso consiste en escoger las variables que queremos ingresar en el modelo. Para esto, lo primero que tenemos que hacer es agregar la variable tiempo, que en este ejemplo se refiere al número de meses que los pacientes con cáncer de pulmón permanecieron con vida. Posteriormente se tiene que definir la variable de resultado en la casilla de Estatus. Habitualmente el evento se codifica con el número 1 (presente). El siguiente paso es seleccionar las covariables que se van a introducir en el modelo, que en este caso son edad, sexo, estadio de la enfermedad y tabaquismo; de estas variables, las últimas tres

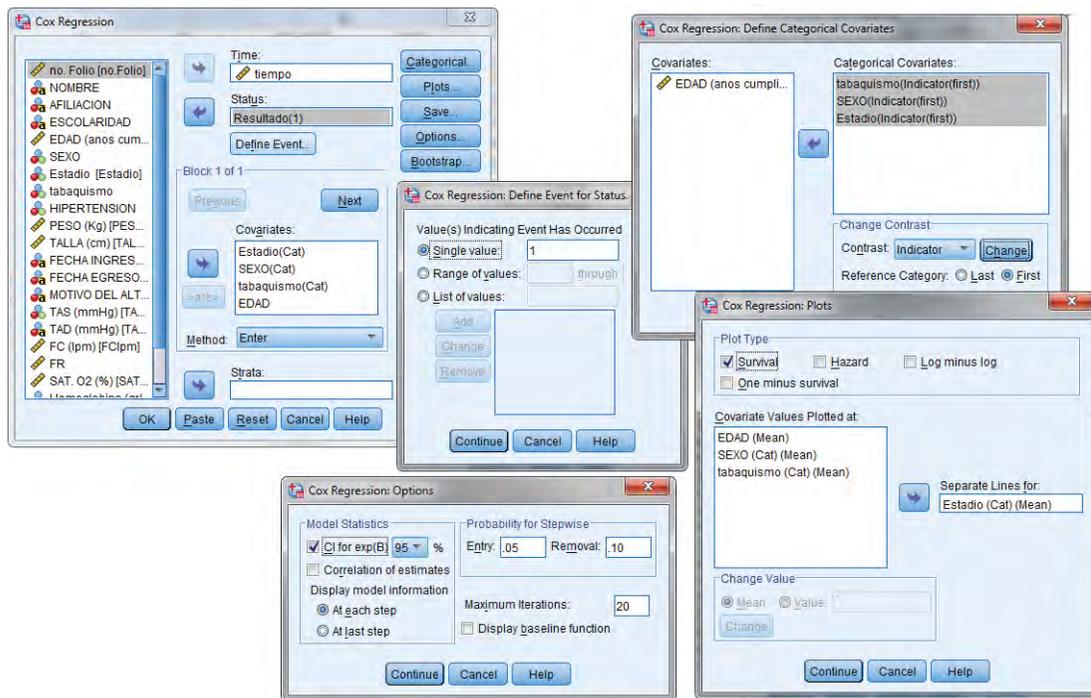


Figura 3 Pasos para la selección de variables categóricas y gráficos en el programa estadístico SPSS para realizar el análisis de riesgos proporcionales de Cox

son categóricas y, por lo tanto, hay que definir las en la opción de Categóricas y determinar cuál de las categorías de cada variable se va a tomar como referente al calcular los riesgos. En este caso nosotros tomamos la primera categoría de referencia, es decir, en la variable sexo las mujeres que están codificadas con el número 1 son el referente, el estadio 1 que está codificado con un 0 será el referente y tabaquismo ausente codificado con un 0 será la referencia. Una vez que ya está definido el evento y seleccionadas las variables que van a entrar al modelo, procedemos a seleccionar los gráficos. Para obtener la curva de supervivencia global se tiene que seleccionar la opción de Supervivencia. Para obtener además las curvas para cada uno

de los estadios de la enfermedad, podemos solicitar que se nos dé la curva con líneas separadas para la o las variables categóricas que deseemos. Esta curva nos dará el resultado de la supervivencia de ese grupo determinado, ajustado por las variables que componen el modelo. Por último, pero no por eso menos importante, debemos pedirle al programa que nos dé los intervalos de confianza al 95 % en el botón Opciones.

La figura 4 muestra la curva de supervivencia global en la que se representa gráficamente cómo van llegando al desenlace las personas conforme pasa el tiempo. Para el mes 60 de seguimiento menos del 40 % de la cohorte continúa con vida. Para la construcción de este gráfico el programa estadístico utiliza

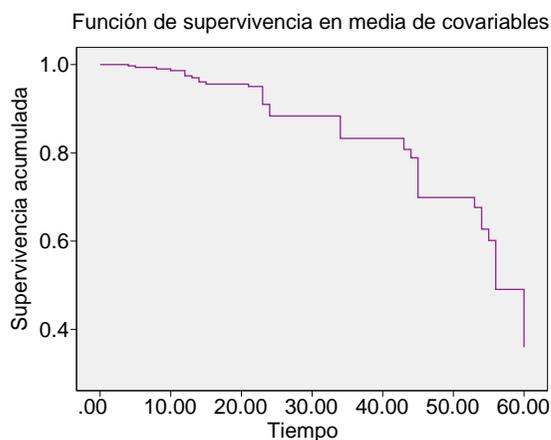


Figura 4 Curva de supervivencia global ajustada por los cofactores

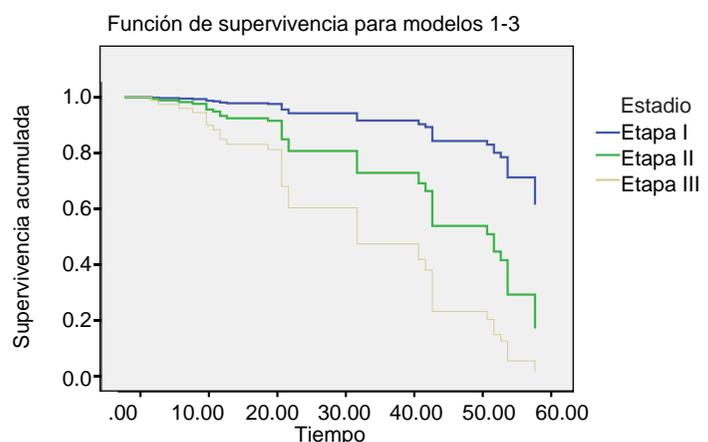


Figura 5 Curva de supervivencia por estadio de la enfermedad

las medias de edad, de estadio de la enfermedad, sexo y tabaquismo como covariables, es decir, no se presentan los riesgos para cada uno de los estadios de la enfermedad.

La figura 5 presenta las curvas de supervivencia ajustadas por edad, sexo y tabaquismo para cada uno de los estadios de la enfermedad, es decir, podemos observar cómo va incrementando la mortalidad en los pacientes en estadio I, II y III de la enfermedad. Es claro que los pacientes en estadio III viven menos que los pacientes en estadios menos avanzados. Para el mes 60 todos los pacientes en el estadio III han presentado el desenlace (muerte), mientras que cerca del 60 % de los pacientes con estadio I continúan vivos.

	Variable en la ecuación					
	B	Wald	Sig	Exp (B)	Inferior	Superior
Edad	.065	31.829	.000	1.067	1.043	1.091
Estadio I		15.843	.000			
Estadio II	1.289	7.230	.007	3.630	1.418	9.291
Estadio III	2.148	15.795	.000	8.568	2.970	24.712
Tabaquismo	.145	.136	.712	1.156	.536	2.492
Sexo (masc)	-.210	.459	.498	.810	.441	1.489

Figura 6 Resultados del modelo de riesgos proporcionales de Cox

La figura 6 muestra los resultados del análisis de Cox. Como en todos los modelos de regresión obtenemos valores de β . En este caso el exponente de β corresponde al HR. En este modelo, la edad y el estadio de la enfermedad son los principales factores de riesgo. La interpretación del HR sería de la siguiente manera: los pacientes en estadio II tienen 3.6 veces el riesgo (o 2.6 veces más riesgo) de haber

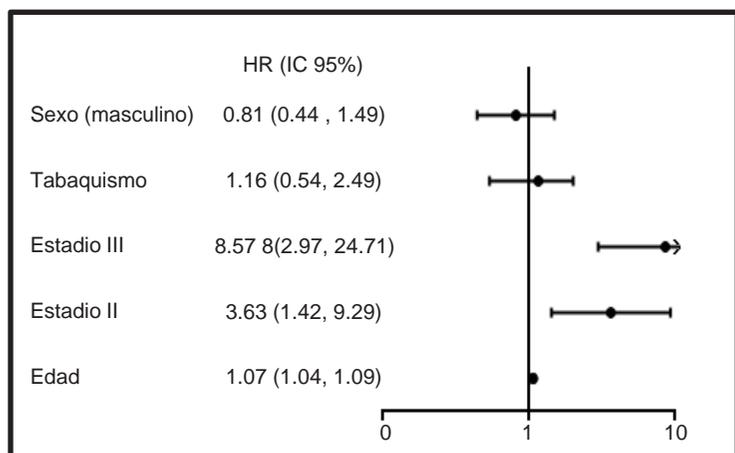
fallecido en el mes 60 de seguimiento en comparación con los que están en estadio I, y los pacientes en estadio III tienen 8.5 veces el riesgo (o 7.5 veces más riesgo) de haber fallecido para el mes 60 de seguimiento en comparación con el estadio I. Cada unidad de la variable edad (cada mes de vida) incrementa el riesgo de muerte en 6.7 %.

Estos HR también se pueden exponer de forma gráfica, usando *forest plots* para mostrar su relación con otras variables, donde la línea vertical en el número 1, es la no asociación (figura 7). Se interpreta de modo similar al OR o al RR. (capítulo VI de esta serie).

Este gráfico permite evaluar la significación estadística a partir de los intervalos de confianza (IC) de 95 %. Cuando el IC 95 % no toca la unidad (línea vertical) representa un estudio que es estadísticamente significativo, como el caso de la edad, el estadio II, y el estadio III. Sin embargo, cuando el IC 95 % toca la unidad e incluso la atraviesa, como en el caso de los pacientes fumadores o el ser hombres, estas variables no tienen relación con el desenlace o no contribuyen con el desenlace, que en este caso es la muerte.

Estos gráficos son muy populares, ya que permiten analizar de un vistazo la relación de las variables y su fuerza de asociación, es decir, no tan solo reflexionar sobre la significación estadística de la relación de las variables tomando en cuenta solo el valor de p , sino lo alejado de la no asociación ($HR = 1$), esto es, el tamaño del efecto, como lo es el estadio III de la enfermedad. Estos gráficos también permiten elaborar nuevas preguntas sobre el tamaño de muestra del estudio, ya que los intervalos de confianza dependen del número de pacientes incluidos, es decir en algunos casos el incrementar el número de observaciones (tamaño de muestra) puede disminuir la amplitud del intervalo de confianza. Dicho de otra manera, con estos gráficos se puede explicar la asociación y el tamaño del efecto de un solo vistazo.

Una vez finalizado el modelo de Cox, como se menciona en el figura 1 debe verificarse su validez y pertinencia en nuestros datos. El primer supuesto tiene que ver con los datos censurados. Debemos asegurarnos de que estos no están relacionados con la probabilidad de que ocurra el evento, es decir, que los sujetos que se pierden en el seguimiento no abandonaron el estudio por razones relacionadas con el tratamiento o la variable de agrupación, como puede ser el estadio de la enfermedad. Este supuesto se verifica investigando en la medida de lo posible las razones de abandono del estudio. Cuando este supuesto se viola, se invalidan todos los análisis de supervivencia desde el modelo de Kaplan-Meier hasta el modelo de Cox.



HR = hazard ratio; IC= intervalo de confianza

Figura 7 Gráfico de las covariantes que afectan la mortalidad por cáncer de pulmón

El segundo supuesto tiene que ver con que los riesgos entre los grupos sean proporcionales. Esto se aplica cuando la variable es cuantitativa y ordinal, y significa que las curvas de supervivencia para cada uno de los estratos deben tener funciones de riesgo que sean proporcionales en el tiempo; por ejemplo, que tengan riesgos con incrementos constantes de un valor a otro (de 50 a 60 años un incremento proporcional similar que de 60 a 70 años). Y que siga la misma dirección, es decir a mayor estadio de la enfermedad, mayor riesgo. Esto puede evaluarse visualmente con un gráfico de log-log (Figura 8) que se solicita en la opción de gráficos en el cuadro de diálogo del modelo de Cox. En nuestro ejemplo, este gráfico muestra que la función de riesgos sí es proporcional entre las etapas de la enfermedad.

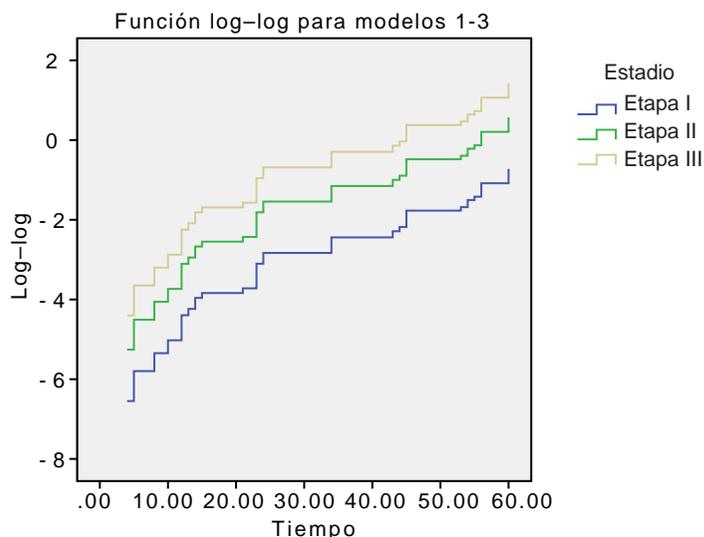


Figura 8 Gráfico de log-log para evaluar función de riesgos proporcionales

Conclusiones

Los modelos de supervivencia son alternativas estadísticas que nos permiten hacer la estimación de un riesgo tomando en cuenta el tiempo transcurrido hasta el desenlace. Este tipo de modelos son bastante útiles en el campo de la salud, especialmente en áreas relacionadas con enfermedades crónicas y enfermedades de alta letalidad. El modelo de riesgos proporcionales de Cox constituye un modelo multivariado que puede ponderar una curva de supervivencia con variables cualitativas o cuantitativas que pueden modificar el efecto final.

Tomar en cuenta todos los posibles factores de riesgo que puedan intervenir en un resultado ayuda a fundamentar el juicio clínico.

Declaración de conflicto de interés: los autores han completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no ha sido reportado alguno que esté relacionado con este artículo.

Referencias

1. Kleinbaum DG, Klein M. *Survival Analysis: A Self-Learning Text*. New York: Springer Science Business Media, Inc; 2005.
2. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 1958;53(282):457-81.
3. Feinstein AR. *Principles of medical statistics*. New York, NY: Chapman and Hall/CRC; 2002.
4. Bradburn MJ, Clark TG, Love SB, Altman DG. Survival analysis part II: Multivariate data analysis—an introduction to concepts and methods. *Br J Cancer*. 2003;89:431-6.
5. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1972;34(2):187-220.
6. Barea-Baltierra R, Rivas-Ruiz R, Pérez-Rodríguez M, Palacios-Cruz L, Moreno J, Talavera JO. Del juicio clínico a la regresión logística múltiple. *Rev Med Inst Mex Seguro Soc*. 2014;52(2):192-7.



Investigación clínica XXIII

Del juicio clínico a los metaanálisis

Rodolfo Rivas-Ruiz^a, Osvaldo D. Castelán-Martínez,^a Marcela Pérez-Rodríguez,^a Lino Palacios-Cruz,^a Maura E Noyola-Castillo,^a Juan O Talavera^a

Yo sabía que no había ninguna evidencia real en el tratamiento que ofrecíamos para la tuberculosis. Me temo que yo acorté la vida de algunos de mis amigos gracias a intervenciones innecesarias.

ARCHIE COCHRANE (1909-1988)

Clinical research XXIII. From clinical judgment to meta-analyses

Systematic reviews (SR) are studies made in order to ask clinical questions based on original articles. Meta-analysis (MTA) is the mathematical analysis of SR. These analyses are divided in two groups, those which evaluate the measured results of quantitative variables (for example, the body mass index —BMI—) and those which evaluate qualitative variables (for example, if a patient is alive or dead, or if he is healing or not). Quantitative variables generally use the mean difference analysis and qualitative variables can be performed using several calculations: odds ratio (OR), relative risk (RR), absolute risk reduction (ARR) and hazard ratio (HR). These analyses are represented through forest plots which allow the evaluation of each individual study, as well as the heterogeneity between studies and the overall effect of the intervention. These analyses are mainly based on Student's *t* test and χ^2 . To take appropriate decisions based on the MTA, it is important to understand the characteristics of statistical methods in order to avoid misinterpretations.

Keywords Palabras clave

Meta-Analysis	Metaanálisis
Review, Systematic	Revisión sistemática
Forest plots	Forest plots

Antecedentes

En las últimas décadas, el avance vertiginoso en las técnicas diagnósticas y en el tratamiento ha llevado a buscar estrategias que ayuden a organizar, revisar, analizar y gestionar de manera eficiente el conocimiento científico, principalmente en beneficio de los pacientes. Este tipo de estrategias han contribuido en los profesionales de la salud (PROSUD) a continuar con su proceso de aprendizaje a lo largo de la vida además de fortalecer su juicio clínico, ya que en muchas ocasiones estos pueden acceder a información de dudosa calidad, a pesar de estar publicada en revistas de prestigio, con un mensaje tan “contundente” que hasta el mejor de los clínicos podría caer en la interpretación incorrecta de los hallazgos.

Usted en este momento podría estarse preguntando si aspectos como la intuición del clínico, las emociones y las motivaciones en la consulta que pueden despertarse en el proceso de la relación médico-paciente, el contexto en el cual se realiza la práctica clínica, la propia experiencia empírica son importantes en los procesos llamados juicio clínico y aprendizaje; sin duda lo son. Sin embargo, para fines del presente artículo el enfoque se dará a una estrategia basada en evidencia científica que tiene que ver con la búsqueda y evaluación sistematizada además del análisis clínico y la toma de decisiones que va de la mano del quehacer del PROSUD. Esta misma estrategia, aplicada de manera correcta, ayudaría a mejorar la comunicación entre el PROSUD y el paciente para saber los alcances de cada opción de diagnóstico y tratamiento, además del pronóstico de cada entidad de acuerdo con las circunstancias para cada caso en particular. Por ejemplo, si se tratase de la fase diagnóstica, el PROSUD podría llegar a recomendar o no la realización de alguna prueba diagnóstica dependiendo de la utilidad o el costo-beneficio que ha mostrado dicha prueba en distintos reportes científicos o en el caso de la fase de tratamiento, o la elección de una u otra opción terapéutica dependiendo de los resultados de estudios que mostrasen su real magnitud de efecto. Finalmente, en el caso del pronóstico se podría emitir una estimación sobre la probabilidad de éxito del tratamiento o la duración de la enfermedad con base en lo que se muestre en estudios clínicos.

^aCentro de Adiestramiento en Investigación Clínica (CAIC), Coordinación de Investigación en Salud, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social

Distrito Federal, México

Comunicación con: Rodolfo Rivas-Ruiz
Correo electrónico: rivasrodolfo@gmail.com

Resumen

Las revisiones sistemáticas (RS) son estudios que intentan contestar una pregunta clínica por medio del uso de artículos originales. Los metaanálisis (MTA) son el análisis matemático de las revisiones sistemáticas. Estos análisis se dividen en dos grandes grupos: aquellos que evalúan los resultados medidos como variables cuantitativas (por ejemplo, el índice de masa corporal —IMC—) y aquellos que evalúan variables cualitativas (por ejemplo, si un paciente está vivo o muerto, o si ha mejorado o no de su enfermedad). Los primeros utilizan en general la diferencia de medias. Para los segundos se puede utilizar el cálculo de

razón de momios (RM) u *odds ratio* (OR), riesgo relativo (RR), reducción absoluta del riesgo (RAR), *hazard ratio* (HR), etcétera. Estos análisis se grafican por medio de *forest plots* que permiten tanto la evaluación individual de los estudios como la del resultado final, así como la heterogeneidad de estas comparaciones. Asimismo, estos análisis se basan en análisis estadísticos básicos como la *t* de Student y la χ^2 . Para la toma de decisiones basadas en las RS y en los MTA, es importante evaluar cuidadosamente los posibles sesgos y los apartados estadísticos con el fin evitar malas interpretaciones.

Ya desde mediados del siglo XIX Bichat y Magendi proponían un cambio sustancial en la manera de establecer las bases clínicas del diagnóstico precoz, el pronóstico y el tratamiento de las enfermedades que podemos considerar como la base de lo que hoy conocemos como la medicina basada en la evidencia (MBE). El término metaanálisis fue empleado por primera vez en 1976 por Gene V. Glass, de la Universidad del Estado de Arizona, para referirse al análisis estadístico del conjunto de resultados obtenidos en diferentes ensayos clínicos sobre una misma pregunta de investigación. Sin embargo, las bases matemáticas se conocían desde el final del siglo XVII cuando Gauss y Laplace lo usaron y describieron estos métodos en el área de la astronomía para disminuir el error.

Las revisiones sistemáticas (RS) y/o metaanálisis (MTA) son compilaciones organizadas y de análisis sistematizado de todos los estudios que hasta ese momento se han realizado sobre algún tema específico, y que tratan de contestar una pregunta de investigación clínica en común. Las RS y MTA han tomado un gran auge en los últimos años, ya que han aparecido como una buena alternativa para los PROSUD, cuyo tiempo para la obtención y el análisis de la evidencia científica suele ser limitado. Sin embargo, estos estudios no están exentos de sesgos que lleven a los clínicos a terminar con una mala interpretación de la información.

Por todos estos motivos, es que consideramos de gran trascendencia el revisar en este número los análisis estadísticos más usados en las RS para realizar MTA y explicaremos el significado de los estadísticos, desde su fundamento hasta los gráficos que se han popularizado en las publicaciones de la colaboración Cochrane y que han permeado incluso en revistas de alto impacto como el *New England Journal of Medicine*.

Empezaremos con un poco de historia. Ya desde mediados del siglo XIX Bichat y Magendi proponían

entusiastamente un cambio sustancial en la manera de establecer las bases clínicas del diagnóstico precoz, el pronóstico y el tratamiento de las enfermedades que podemos considerar como la base de lo que hoy conocemos como la medicina basada en la evidencia (MBE).

Hoy en día, los conceptos han cambiado y se conciben las RS y los MTA como estudios secundarios. Esto significa que compilan los resultados de otros estudios con diversos fines. Entre los más destacados tenemos:

1. Ponderar los resultados obtenidos en distintos artículos del mismo tema. A esta circunstancia también se le ha llamado validez externa, pues se trata de una evaluación de los artículos realizada por otros autores diferentes a los autores que hicieron los artículos originales. Esto supone una revisión de los posibles sesgos que concluya en un resultado final con mayor credibilidad, que sea transparente y directo al resolver una pregunta clínica específica en un solo documento, a fin de que el clínico, tras usar su juicio, logre tomar la mejor decisión ante el paciente que tiene delante de él.
2. Evaluar la consistencia entre los ensayos clínicos de similares intervenciones inherentes al tema y generar un estimador del efecto más eficiente.
3. Mejorar el poder estadístico, es decir, la precisión de los resultados arrojados por artículos vistos de manera individual que por tener un tamaño de muestra pequeño no logren tener resultados estadísticamente significativos.
4. Obtener información para el cálculo del tamaño de muestra para estudios futuros.

Cuando las RS tienen valores cuantitativos (como la media del índice de masa corporal —IMC—) o cualitativos dicotómicos (vivo o muerto) y estos son susceptibles de ser comparados y sumados, a estos análisis se les llama metaanálisis. Es decir los MTA son la parte matemática de las RS. Es común que se

piense que es mejor un MTA que una RS, cuando en la realidad el mecanismo de búsqueda de la información es el mismo y solo difieren en la parte matemática; de hecho, muchas veces cuando el clínico no está adiestrado en la lectura de los MTA puede ser fácilmente manipulado por autores que se valen de herramientas estadísticas para mostrar resultados que no son necesariamente ciertos o que no tienen una relevancia clínica.

En general, tanto las RS como los MTA requieren de una metodología llamada sistema PICO o PICOST,¹ la cual permite hacer explícitos los métodos que llevaron a localizar los artículos.

En este sentido las revisiones sistemáticas con MTA usan distintas técnicas estadísticas para intentar analizar cada uno de los resultados de un artículo y luego compilarlos y resumirlos para dar un resultado final. En específico, en los MTA se pretende contestar preguntas en las que tenemos un estado basal similar entre estudios. Este se verifica mediante criterios de inclusión de los artículos. Casi siempre se comparan únicamente dos intervenciones, lo que se denomina como “cara a cara”. Estos estudios son analizados de manera individual (para reproducir los resultados de los autores) y finalmente son compilados en un resultado final (en un rombo final).

El gráfico más popular de los metaanálisis es el llamado *forest plot*, que recibe ese nombre por su apariencia similar a la de un bosque. En los *forest plots* la línea vertical del gráfico señala la ausencia de efecto y se despliega el efecto para favorecer a una u otra maniobra. Esta siempre se muestra en los gráficos, en el área de las columnas. El resultado se muestra en el área del gráfico (figuras 1a y 1b).

En este ejemplo, el gráfico muestra que el efecto del tratamiento favorece el uso de la aspirina al compararlo con el placebo para el manejo de una cefalea. En el ejemplo de consumo de café frente al del placebo no hay diferencias como antioxidante y en el ejemplo del grupo de sujetos con dislipidemias se favorece al grupo de los sanos (sin dislipidemia) para la prevención de infartos agudos al miocardio.

Estos gráficos son usados para evaluar tanto resultados cualitativos dicotómicos (como puede ser el estar vivo o muerto, o el estar infartado o sin infarto en pacientes usuarios de aspirina profiláctica) como resultados cuantitativos (horas de sueño posteriores a la ingesta de café o placebo). Cada uno de los artículos se representa con su comparación real (tal cual está publicada), que se basa en los números crudos y se representa gráficamente como un cuadrado, que es la media del efecto, y por intervalos de confianza (en general del 95 %) que se representan con una línea horizontal que atraviesa el cuadrado.

Diferencia de medias

Cuando el desenlace es cuantitativo, el gráfico se llama *forest plot* de diferencia de medias. Este análisis se basa en los estadísticos de la *t* de Student para grupos independientes, con la diferencia de que en este caso el MTA es la suma de todos los resultados de cada uno de los artículos que forman parte de la revisión sistemática.

A continuación presentamos la fórmula general del MTA de la diferencia de medias, donde δ es igual al resultado final del diamante final (del *forest plot*) de

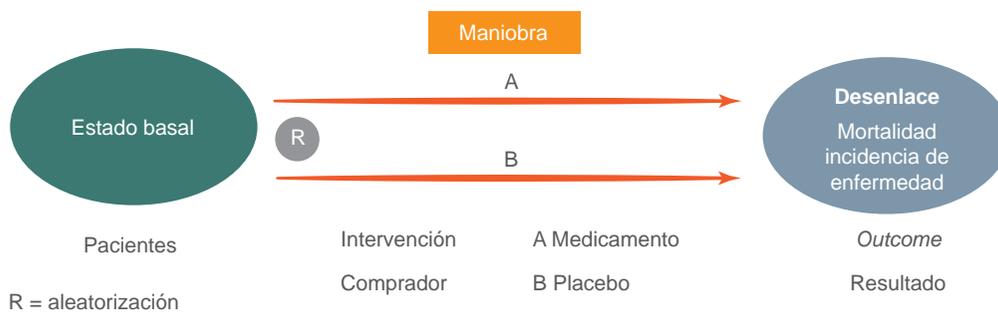


Figura 1a Ejemplo de las partes de la arquitectura de la investigación cuyas equivalencias las encontrará resaltadas el lector con colores en la figura 1b

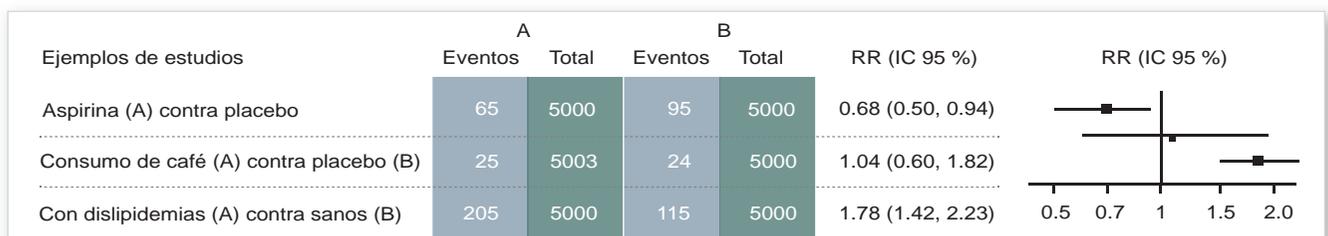


Figura 1b Partes de un *forest plot*

Favorece A Favorece B

los MTA. Y este resultado es igual a la suma de la media de cada uno de los estudios del grupo de tratamiento activo y se le resta la misma medida de los del grupo control o de comparación, todo lo anterior dividido entre la desviación estándar:

$$\delta = \frac{\mu_t - \mu_c}{\sigma}$$

El análisis de cada uno de los artículos se realiza de modo similar a la *t* de Student. En un ejemplo hipotético (figura 2) en el que se estudian dos grupos de veinte pacientes y en el que se compara un medicamento para bajar de peso (A) con un placebo (B), el primer análisis consistiría en la comparación del IMC entre los grupos por medio de la prueba *t* de Student para grupos independientes (como vimos en Investigación clínica XV).²

En la figura 3 podemos ver que el análisis individual de los artículos es similar al que se realiza con la *t* de Student. Sin embargo, el *forest plot* tiene la ventaja de mostrar la diferencia de medias (en verde) y los intervalos de confianza del 95 % (IC 95 %).

Como vimos en el artículo de Investigación clínica VI (cuyo tema es la relevancia clínica),³ en el que mencionamos la importancia del uso de los IC del 95 %, en especial se puede obviar el valor de *p*, ya que si los intervalos de confianza atraviesan la unidad, esta comparación no es estadísticamente significativa.³ En este ejemplo podemos ver las partes de un *forest plot*. En la columna donde dice Weight se hace referencia al peso de cada artículo sobre el resultado final que se aprecia en forma de rombo negro. En este caso hipotético solo hay un artículo, por lo que se

infiere que el resultado final se debe en un 100 % al artículo de Rivas-Ruiz del 2013. En las filas podemos observar además los valores crudos del estudio con su media y su desviación estándar; una fila más abajo, la sumatoria total con su resultado. Y más abajo se presenta el análisis de heterogeneidad entre los estudios. Como solo se trata de un estudio, no es aplicable llevar a cabo este análisis. El tema de la heterogeneidad lo vamos a abordar más adelante.

En la línea final se encuentra el análisis estadístico para obtener el valor de *p* (la probabilidad de que el efecto final del metaanálisis —rombo negro— se deba al azar). Al igual que todos los análisis, este valor de *p* debe ser < 0.05 para considerarse significativo. En este caso el valor es de 0.06, lo que refiere que los tratamientos de A y B no son distintos en la efectividad para disminuir el IMC.

Ahora realizaremos combinaciones hipotéticas para ejemplificar con los componentes de las RS y los MTA.

Siguiendo el ejemplo hipotético, si combináramos los resultados de un estudio con otro estudio con menor tamaño de muestra, para incrementar el poder estadístico, pero con resultados similares (figura 4), lo que se encontraría es que el resultado final (rombo negro) tendría ya un resultado estadísticamente significativo; en este caso a favor del tratamiento B para la reducción del IMC.

Siguiendo el orden anterior, tendríamos el resultado de la *t* de Student del segundo estudio hipotético (Pérez-Rodríguez, 2014), el cual no resultó significativo *p* = 0.08. Sin embargo, al combinar los dos estudios el resultado final (rombo negro) resulta a favor del tratamiento B (*p* = 0.02).

Como ya se había mencionado, la columna del apartado de peso (Weight) hace referencia al peso que

Maniobra		N	Media	Des. Est	Error típico de la media
IMC	Tx A	20	34.2000	1.70448	.38113
	Tx B	20	32.8500	2.70039	.60383

Resultado de la *t* de Student de la comparación hipotética de la maniobra A frente a la B

		Prueba de Levene para la igualdad de varianzas		Prueba T para la igualdad de medias						
		F	Sig.	t	gl	Sig. (bilateral)	Diferencia de medias	Error típico de la diferencia	95 % intervalo de confianza para la diferencia	
								Inferior		Superior
IMC	Se han asumido varianzas iguales	2.554	.118	1.891	38	0.66	1.35000	.71405	-.09552	2.79552
	No se han asumido varianzas iguales			1.891	32.066	0.68	1.35000	.71405	-.10436	2.80436

Figura 2 Análisis de la maniobra A frente a la B mediante la prueba *t* de Student

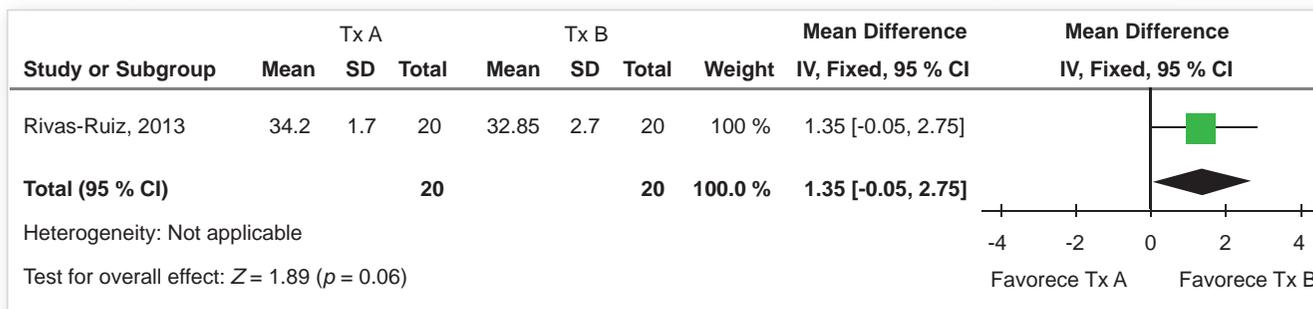


Figura 3 Forest plot del análisis individual de la comparación hipotética de A frente a B

tiene cada artículo a partir de la cantidad de pacientes que tiene cada estudio. La prueba de heterogeneidad de χ^2 que se observa en la figura evalúa si los resultados son homogéneos o heterogéneos: cuando el valor de p de la prueba es < 0.05 se dice que los resultados son heterogéneos y cuando el valor de p es ≥ 0.05 la prueba resulta no significativa, lo que quiere decir que los estudios tienen resultados homogéneos (es decir, similares, en este caso idénticos). Las siglas df de la prueba de heterogeneidad hacen referencia a los grados de libertad que se calculan restando una unidad al número de estudios incluidos en el análisis; en este caso, en el que son dos estudios los incluidos, se tiene 1 grado de libertad ($2 \text{ estudios} - 1 = 1$).

Siguiendo con el ejemplo hipotético, si añadiríamos otro estudio publicado con anterioridad con resultados completamente contrarios, el efecto final estadísticamente significativo desaparecería y refutaría los hallazgos de los otros dos artículos, lo cual nos daría como resultado que no hay diferencia entre ambos tratamientos ($p = 1.00$). En la figura 5, podemos observar que al incorporar un artículo más con mayor número de pacientes la columna de peso vuelve a cambiar y le da mayor peso (50 %) al grupo con más pacientes. Además, en el análisis de heterogeneidad la χ^2 con dos grados de libertad (df) por tener tres estudios tiene un valor de $p = 0.005$, lo cual significa que alguno de los estudios tiene resultados heterogéneos (en este ejemplo el heterogéneo es el estudio del doc-

tor Talavera, el cual es completamente contrario a los dos artículos publicados tiempo después).

Cuando los resultados de los estudios son heterogéneos se recomienda hacer el análisis de forma más robusta con efectos aleatorios, lo cual se explicará con más detalle más adelante. Con esto en mente encontramos que debido a que el estudio del doctor Talavera incorporó más pacientes y tiene un mayor peso, la heterogeneidad medida con I^2 es del 81 %.

Debido a que la comparación presentó esta heterogeneidad alta, el MTA debe ser analizado por efectos aleatorios (random), lo cual vemos en la parte superior de la figura 6.

Cuando se analiza por métodos aleatorios, no se toma en cuenta solo el peso de cada uno de los artículos, sino el del efecto y el de los intervalos de confianza. Este análisis cambia el resultado final, como podemos ver en la fila del total (95 % CI), el cual a diferencia de la figura 5, donde la diferencia era de 0, ahora al cambiar el criterio de ponderación, presenta una tendencia ligeramente a favor del Tx B con una diferencia de 0.37 de IMC, sin ser estos estadísticamente significativos ($p = 0.72$).

Diferencia de proporciones

Este es uno de los análisis más populares, ya que en la práctica clínica la pregunta directa de los pacientes es:

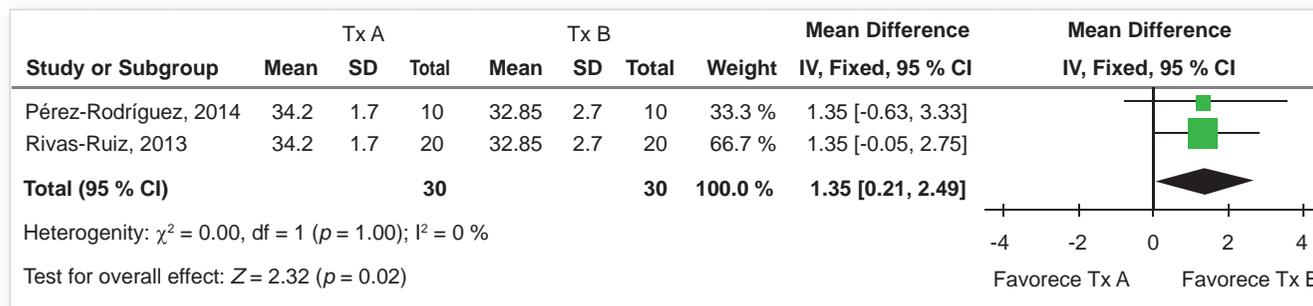


Figura 4 Forest plot del análisis de dos artículos de la comparación hipotética de A frente a B



Figura 5 Forest plot de la comparación de tres artículos con dos maniobras con un resultado cuantitativo, con heterogeneidad, analizado por métodos fijos

“¿Qué medicamento es mejor A o B?”, por lo que este análisis es uno de los más utilizados.

Se utiliza el *forest plot* para la diferencia de proporciones cuando el resultado del estudio es dicotómico (por ejemplo, enfermo o no enfermo). Hoy en día el *software* Rev Man (disponible en <http://tech.cochrane.org/revman/download>) permite realizar distintos análisis y se puede calcular razón de momios (RM) o *odds ratio* (OR) cuando se comparan estudios transversales y casos y controles. En el caso de cohortes y ensayos clínicos, se calculan los riesgos relativos (RR). También se pueden hacer cálculos usando la reducción absoluta del riesgo (RAR). Recientemente se han incorporado análisis para estudios de seguimiento, en los que la medida de asociación es el *hazard ratio* (HR), el cual es producto de estudios de supervivencia.

El método estadístico que usan estos análisis tiene su base estadística en la χ^2 , por medio de la cual se considera lo observado frente a lo esperado.⁴

Estos *forest plots* tienen las mismas características que las enunciadas previamente, en las que a cada uno de los estudios se le analiza de manera individual en cada renglón con la fórmula que le corresponda (RM, RR o HR).³

Siguiendo con el ejemplo hipotético, en el cual se busca cuál es la eficacia de un tratamiento A al compararlo con un tratamiento B, y en el que ahora el resul-

tado es que los pacientes lleguen a un peso adecuado, tendríamos lo que se muestra en la figura 7.

En estos ejemplos encontramos que los estudios de Pérez-Rodríguez y Rivas-Ruiz muestran que un mayor número de pacientes en el tratamiento B llegan a un peso adecuado en mayor porcentaje que los pacientes que tomaron el tratamiento A, a diferencia del estudio de Talavera, el cual refiere que no hay diferencias significativas entre los dos grupos y se representa gráficamente con los IC 95 % que tocan la unidad.

Sin embargo, el resultado final (rombo negro) nos hace referencia a que el efecto general de la comparación A frente a B favorece a la administración del tratamiento B y además esto resulta estadísticamente significativo ($p = 0.009$).

En la figura 7, presentamos todas las posibilidades de análisis. Sin embargo, lo correcto sería un análisis con RR (B), debido a que los estudios fueron ensayos clínicos.

Heterogeneidad

Como vimos en los ejemplos anteriores, se utilizan dos métodos para demostrar la homogeneidad o la heterogeneidad de los resultados que se incorporan en un MTA. El primero consiste en una prueba de hipótesis, con la cual se busca conocer si existe o no

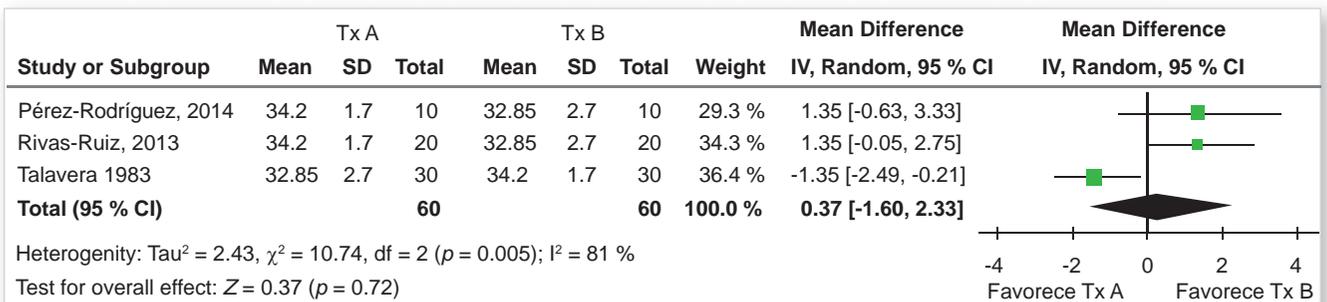


Figura 6 Forest plot de la comparación de dos maniobras con un resultado cuantitativo, con heterogeneidad, analizado por métodos aleatorios

heterogeneidad. Esa prueba se basa en la siguiente prueba de hipótesis:

$H_0 = E_1 = E_2$ (los resultados de los estudios [E] son homogéneos)

$H_a = E_1 \neq E_2$ (los resultados de los estudios [E] son heterogéneos)

Para realizar esta prueba, se utiliza en general la prueba de χ^2 . El valor crítico para desechar la hipótesis nula es un valor de $p < 0.05$.

Sin embargo, si queremos conocer el grado de heterogeneidad, se requiere de otro estadístico llamado I^2 , el cual se describe en forma de porcentaje, y va del 0 al 100 %, donde el 0 es la heterogeneidad nula (total homogeneidad) y el 100 % es completamente heterogéneo. En general el punto crítico de la I^2 es el 50 %. Sin embargo este valor puede ser modificado a consideración de los investigadores. En los ejemplos anteriores de la figura 7 encontramos que en todos los casos la I^2 demostraba que los

resultados eran heterogéneos, es decir, dos artículos refieren que el fármaco B es mejor que el A y un artículo lo contrario, por lo que los resultados son contradictorios.

Con estas pruebas se pueden tomar decisiones para realizar los análisis del MTA, ya sea por métodos fijos (en caso de que los estudios sean homogéneos) o por métodos aleatorios (en caso de que los estudios sean heterogéneos).

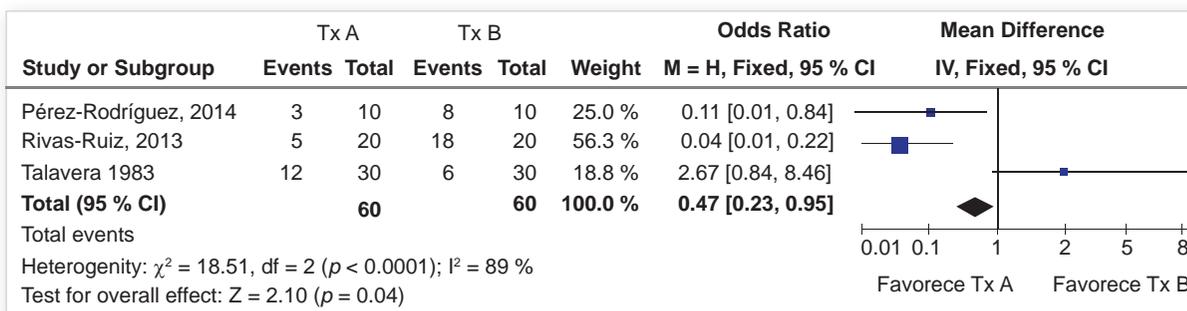
Los métodos aleatorios son más robustos que los métodos fijos, por lo que se recomiendan cuando los resultados no van a la misma dirección.

Regresando a nuestro caso hipotético, encontramos que la heterogeneidad es significativa ($p < 0.001$) y que la I^2 siempre es > 50 %, por lo que el análisis más conveniente debe de ser por efectos aleatorios (figura 8).

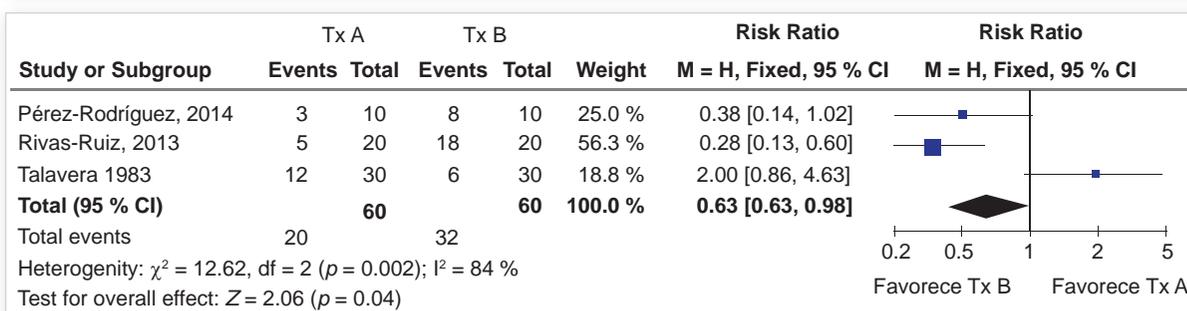
Como se puede apreciar, ahora el resultado del MTA es no significativo ($p = 0.42$) a pesar de que parece favorecer al tratamiento B.

Por lo anterior, es muy importante seleccionar la prueba adecuada para el análisis de estos datos, ya que

(A)
Analizado con RM por efectos fijos



(B)
Analizado con RR por efectos fijos



(C)
Analizado con RAR por efectos fijos.

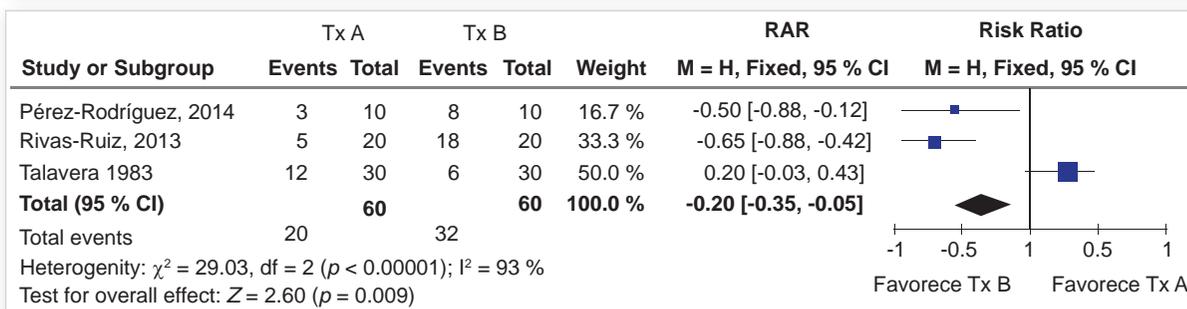


Figura 7 En el forest plot (A) se hace un análisis de los tres estudios con RM (B), después con RR (C) y finalmente con reducción absoluta del riesgo (RAR)

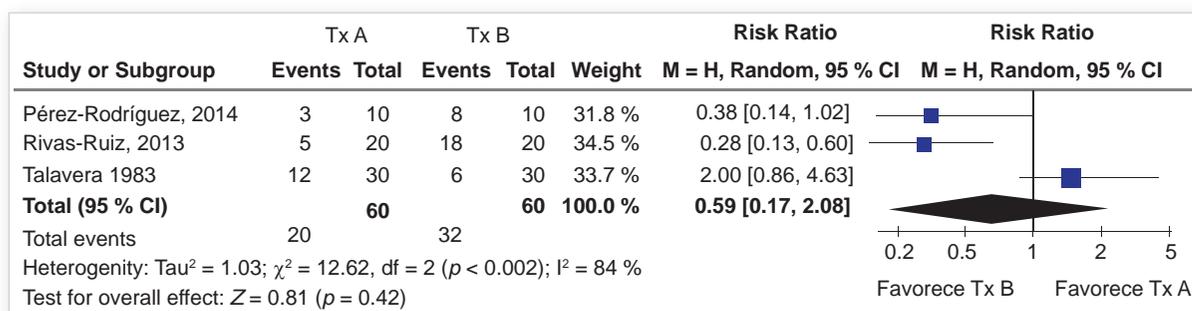


Figura 8 Forest plot de la comparación de A frente a B mediante efectos aleatorios

a partir de la selección del análisis los resultados pueden resultar contrarios.

A pesar de que las fórmulas para los MTA son muy sencillas, hoy en día se requiere que estos análisis se realicen con *software* especializado. El *software* que nosotros recomendamos es el que distribuye de manera gratuita la colaboración Cochrane, el cual se llama Rev Man y, como mencionamos antes, puede ser descargado directamente de su página de Internet. También se pueden utilizar otros programas estadísticos como STATA.

Conclusiones

Los MTA son la ponderación matemática de los estudios cuantitativos. Cuando la variable dependiente es una variable cuantitativa (como peso, talla o IMC), entonces se utiliza la diferencia de medias.

Cuando la variable dependiente es dicotómica (variable cualitativa, como vivo o muerto, mejoría o no mejoría), se pueden utilizar distintos modelos dependiendo del tipo de diseño de los artículos originales (puede ser RM, RR o HR).

Los MTA son de utilidad por compilar la información de una revisión sistemática. Los estadísticos pretenden ser sencillos y son graficados para que puedan ser leídos a primera vista. Sin embargo, se debe tener cuidado en la interpretación y el uso adecuado de los modelos estadísticos.

Algunos investigadores y personal de la salud están en contra de este tipo de estudios, al referir que

al juntar basura solo se obtiene más basura. Existe controversia también sobre el peso de los MTA en los niveles de evidencia que utilizan las guías de práctica clínica. Esta controversia radica en que es preferible usar la información de un solo ensayo clínico aleatorizado multicéntrico, la cual otorga una información más robusta al venir de un solo protocolo.

Sin embargo, el principio de las RS y los MTA es tener un protocolo lo suficientemente limitado que permita combinar pacientes y maniobras similares. Esto se consigue cuando elaboramos la pregunta de investigación utilizando la palabra PICO o PICOST, con lo cual delimitamos la población blanco, la intervención, el comparador, los resultados, y otras características particulares, como el tipo de estudio que nos permitirá elegir aquellas publicaciones que respondan de forma clara y específica la pregunta planteada. Por eso creemos que entender la estadística de este tipo de estudios ayudará sin duda a modular el juicio clínico y a mejorar la toma de decisiones en la práctica cotidiana.

También consideramos que todos los estudios están sujetos a distintos sesgos y por lo tanto entenderlos y analizarlos mejor ayuda a desmitificar este tipo de controversias y a valorar los estudios en su justa dimensión.

Declaración de conflicto de interés: los autores han completado y enviado la forma traducida al español de la declaración de conflictos potenciales de interés del Comité Internacional de Editores de Revistas Médicas, y no fue reportado alguno en relación con este artículo.

Referencias

- Rivas-Ruiz R, Talavera JO. [Clinical Research VII]. Systematic search: how to look for medical documents]. Rev Med Inst Mex Seguro Soc. 2012 Jan-Feb;50(1):53-8.
- Rivas-Ruiz R, Pérez-Rodríguez M, Talavera JO. [Clinical research XV. From the clinical judgment to the statistical model. Difference between means. Student's t test]. Rev Med Inst Mex Seguro Soc. 2013 May-Jun;51(3):300-3.
- Talavera JO, Rivas-Ruiz R. [Clinical research VI. Clinical relevance]. Rev Med Inst Mex Seguro Soc. 2011 Nov-Dec;49(6):631-5.
- Rivas-Ruiz R, Castelán-Martínez OD, Pérez M, Talavera JO. [Clinical research XVII. χ^2 test, from the expected to the observed]. Rev Med Inst Mex Seguro Soc. 2013;51(5):552-7.