



**6**

## **PRINCIPALES TÉCNICAS ESTADÍSTICAS Y PARÁMETROS PARA MEDIR RESULTADOS EN SALUD**

FERNANDO RODRÍGUEZ ARTALEJO, TERESA VILLAGARCÍA CASLA,  
ANA ORTEGA ESLAVA



## ■ 6.1. MEDIDAS DE ASOCIACIÓN O EFECTO EN ESTUDIOS EPIDEMIOLÓGICOS

### ■ 6.1.1. Introducción

Habitualmente el efecto de un factor de riesgo sobre una enfermedad se evalúa en estudios epidemiológicos, generalmente de cohortes y de casos-controles, y el efecto de una intervención sanitaria sobre una enfermedad se suele evaluar mediante ensayos clínicos. Un estudio de cohortes es un estudio epidemiológico en el que se selecciona a sujetos libres de la enfermedad o fenómeno de salud que se investiga y se mide la exposición a sus potenciales determinantes. Se les sigue a lo largo del tiempo y se compara la frecuencia con que aparece (incidencia) el fenómeno de salud en estudio entre los expuestos y no expuestos a cada uno de los potenciales determinantes. Un ensayo clínico es un diseño similar en el que la exposición (e.g. una intervención sanitaria) se asigna (habitualmente al azar) por el investigador a los grupos del estudio.

Hay dos formas de medir la incidencia (aparición de casos nuevos) de la enfermedad o variable resultado en los estudios de cohortes y en los ensayos clínicos. La primera es la incidencia acumulada, en la que se divide el número de sujetos en los que aparece la enfermedad a lo largo del seguimiento por el número de sujetos libres de la misma al inicio del seguimiento; por tanto, la incidencia acumulada es una proporción. La segunda forma es la tasa de incidencia, en la que se divide el número de sujetos en los que aparece la enfermedad a lo largo del seguimiento por la suma de los tiempos en que cada sujeto ha sido seguido; por tanto, es una tasa (enfermedad/tiempo de seguimiento). El análisis de los datos consiste en calcular la incidencia de la variable resultado (enfermedad, muerte, etc.) y compararla entre los sujetos expuestos y los no expuestos al potencial factor de riesgo/intervención sanitaria. Cuando la mayoría de los sujetos se han seguido aproximadamente durante el mismo tiempo, la medida más adecuada de incidencia es la llamada incidencia acumulada (riesgo). Si no es así, debe calcularse la densidad de incidencia.

Las diferencias absolutas o relativas (razones) de las incidencias de la enfermedad en los dos grupos de comparación constituyen medidas de la asociación entre el factor de riesgo/intervención sanitaria y la enfermedad; si se considera que

la relación es causal, constituyen medidas del efecto del factor de riesgo/intervención sanitaria. Estas medidas se calculan de la misma forma y tienen el mismo significado para la incidencia acumulada y para la densidad de incidencia. Por ello, realizaremos los cálculos usando incidencias acumuladas.

### ■ 6.1.2. El riesgo relativo y la *odds ratio*

El riesgo relativo (RR) es la incidencia de enfermedad en los expuestos dividida por la incidencia en los no expuestos al factor de riesgo (Tabla 1). Es una medida de la fuerza de la asociación, y proporciona información útil para juzgar si una asociación, que no se debe al azar ni a errores sistemáticos, puede ser causal; por ello es la forma más habitual de presentar los resultados de los estudios etiológicos. Carece de unidades, su límite inferior es 0 y el superior es infinito. Un RR=1 significa que ambas incidencias son iguales y, por tanto, el potencial factor de

Tabla 1.

#### MEDIDAS DE ASOCIACIÓN Y EFECTO EN ESTUDIOS DE COHORTES

Medida	Significado	Cálculo
Riesgo relativo (razón de riesgos)	Cuantas veces es más frecuente la incidencia de la enfermedad en los expuestos que en los no expuestos al factor de riesgo.	$RR = I_e / I_o$
Riesgo atribuible (diferencia de riesgos)	Incidencia de la enfermedad que es atribuible a la exposición al factor de riesgo.	$RA = I_e - I_o$
Proporción de riesgo atribuible	Proporción de la enfermedad que es atribuible a la exposición al factor de riesgo.	$RA \% = (I_e - I_o) / I_e \times 100$
Índice de intervención	Número de sujetos en los que hay que impedir la exposición para evitar una enfermedad.	$II = 1 / RA$
Riesgo atribuible poblacional	Incidencia de la enfermedad que es atribuible a la exposición al factor de riesgo en la población.	$RAP = I_t - I_o = RA \times P_e$
Proporción de riesgo atribuible poblacional	Proporción de la enfermedad que es atribuible a la exposición al factor de riesgo en la población.	$RAP \% = (I_t - I_o) / I_t \times 100$

*Nota:*  $I_e$ : incidencia de enfermedad entre los expuestos.  $I_o$ : Incidencia de enfermedad entre los no expuestos.  $I_t$ : Incidencia de la enfermedad en el conjunto de la población.  $I_t$ : Incidencia total de la enfermedad en el conjunto de la población.  $P_e$ : Prevalencia de la exposición al factor de riesgo en el conjunto de la población.

*Fuente:* Elaboración propia.

riesgo carece de efecto sobre la aparición de la enfermedad. Valores superiores a 1 significan que el factor aumenta el riesgo de enfermar. La mayoría de los factores de riesgo toman valores entre 1 y 2. Se trata de efectos pequeños o moderados que requieren muestras de gran tamaño para alcanzar la significación estadística, es decir, para demostrar que son diferentes a 1. En general, hay que desconfiar de valores muy altos del RR. Suelen deberse a errores de cálculo o son hallazgos obvios. Por ejemplo, obtener que el RR de muerte por desplazarse en automóvil a 200 km/h es 25 en relación a los que conducen a 90 km/h resulta una obviedad. Sin embargo, también hay ejemplos en los que no se cumple esta regla. Así el tabaco, que es el factor aislado responsable de más muertes en las sociedades occidentalizadas y que presenta una gran prevalencia, tiene RR superiores a 5 para el cáncer de pulmón. Otros factores también pueden presentarlos, como el asbesto en la producción de mesotelioma y cáncer de pulmón, pero su prevalencia en la población general es escasa. Un RR inferior a 1 significa que el factor de estudio reduce el riesgo de la variable resultado (enfermar, morir, etc.); no se trata de un factor de riesgo sino de un factor protector. Es el resultado que se espera obtener cuando se evalúa un medicamento o una vacuna.

Cuando la variable de resultado es la mortalidad general y se usa la incidencia acumulada, es necesario acompañar el valor del RR del tiempo de seguimiento de los participantes en el estudio, pues el RR se irá aproximando a 1 en el transcurso de los años según las incidencias se aproximen al 100 % en los dos grupos

Tabla 2.

**MEDIDAS DE ASOCIACIÓN ENTRE EL CONSUMO DE CIGARRILLOS Y EL CÁNCER DE PULMÓN EN VARONES\***

Mortalidad por cáncer de pulmón entre fumadores de cigarrillos:	209/100.000/año
Mortalidad por cáncer de pulmón entre <b>no</b> fumadores de cigarrillos:	14/100.000/año
Mortalidad por cáncer de pulmón en el conjunto de la cohorte (fumadores y no fumadores):	141/100.000/año
Prevalencia de consumo de tabaco en el conjunto de la cohorte (fumadores y no fumadores):	0,65 (65 %)
<b>Medidas de asociación:</b>	
<b>Riesgo relativo:</b> $(209/100.000) / (14/100.000) = 14,9$	
<b>Riesgo atribuible:</b> $(209/100.000) - (14/100.000) = 195/100.000$	
<b>Proporción de riesgo atribuible:</b> $(195/100.000) / (209/100.000) \times 100 = 93,3 \%$	
<b>Índice de intervención:</b> $1 / (195/100.000) = 512$	
<b>Riesgo atribuible poblacional:</b> $(141/100.000) - (14/100.000) = (195/100.000) \times 0,65 = 127/100.00$	
<b>Proporción de riesgo atribuible poblacional:</b> $(127/100.000) / (141/100.000) \times 100 = 90,1 \%$	

Nota: \*Estimado a partir de Doll *et al.*, 1994.

Fuente: Baker (2016).

de comparación. En cualquier caso, siempre es una buena práctica, ya que el RR puede cambiar durante el seguimiento, según se va manifestando el efecto del factor riesgo al alcanzarse el tiempo de inducción o a medida que envejecen los miembros de la cohorte. Por ejemplo, algunos de los factores de riesgo clásicos de las enfermedades cardiovasculares, como la colesterolemia o el tabaco, reducen su valor en los ancianos.

La Tabla 2 presenta las medidas de asociación entre el consumo de tabaco y la aparición de cáncer de pulmón calculadas a partir del seguimiento de 40 años de la cohorte de médicos británicos (Baker, 2016). Es un estudio diseñado para establecer si el consumo de tabaco aumentaba el riesgo de cáncer de pulmón, y para ello se comparó la incidencia de cáncer entre fumadores y no fumadores. En este caso un RR de 14,9 significa que, en un período de 40 años, los fumadores tienen 14,9 veces más riesgo de morir por cáncer de pulmón que los no fumadores.

El RR también puede calcularse cuando la variable de exposición es continua. Generalmente se buscan puntos de corte que definen categorías de exposición, y se calcula el RR de cada categoría comparando la incidencia de enfermedad en cada una de ellas con la de referencia, que suele ser el nivel más bajo de exposición (Tabla 3). Los puntos de corte han de establecerse en el diseño del estudio y suelen atender a uno de los siguientes criterios: a) grupos de exposición de igual tamaño definidos, por ejemplo, por cuartiles o quintiles del factor de riesgo, y b) grupos definidos por la magnitud de la exposición y su significado biológico o conductual. En el caso del tabaco, los puntos de corte pueden ser no fumadores, fumadores de 1-14 cigarrillos, de 15-24, y de 25 o más. Una vez calculados los RR para cada nivel de exposición, puede estudiarse si existe una relación dosis-respuesta y la función matemática que mejor la describe.

Tabla 3.

### CÁLCULO DE LOS RIESGOS RELATIVOS EN UN ESTUDIO DE COHORTES SOBRE EL CONSUMO DE TABACO Y EL CÁNCER DE PULMÓN CON VARIAS CATEGORÍAS DE EXPOSICIÓN

	Mortalidad por cáncer de pulmón
No fumadores	14/100.000/año
Fumadores de 1-14 cigarrillos	105/100.000/año
Fumadores de 15-24 cigarrillos	208/100.000/año
Fumadores de $\geq 25$ cigarrillos	355/100.000/año
<b>*RR de fumadores de 1-14 cigarrillos frente a no fumadores:</b>	$(105/100.000) / (14/100.000) = 7,5$
<b>RR de fumadores de 15-24 cigarrillos frente a no fumadores:</b>	$(208/100.000) / (14/100.000) = 14,9$
<b>RR de fumadores de <math>\geq 25</math> cigarrillos frente a no fumadores:</b>	$(355/100.00) / (14/100.000) = 25,4$

Notas: \*RR: Riesgo relativo.

Fuente: Datos tomados de Doll *et al.* (1994).

Otra medida de asociación utilizada frecuentemente es la *odds ratio*; aunque se calcula de forma algo diferente a un RR, lo importante es que se interpreta de forma similar. La *odds ratio* se puede obtener en estudios transversales, de casos-controles y de cohortes, y se ha popularizado mucho porque los modelos de regresión logística, que son muy usados en el análisis de estudios epidemiológicos, proporcionan *odds ratios* en lugar de RR.

### ■ 6.1.3. El riesgo atribuible y medidas relacionadas

El riesgo atribuible (RA) es la diferencia de la incidencia de enfermedad entre los expuestos y no expuestos al factor de riesgo (Tabla 1). Es una medida del efecto absoluto del factor de riesgo y representa el “exceso” de riesgo de enfermar, entre los expuestos, atribuible al factor de riesgo. Por tanto, a diferencia de lo que ocurre con el RR, el cálculo del RA implica la asunción de una relación causal entre el factor de riesgo y la enfermedad. El límite inferior del RA es 0 y el superior es 100 o infinito, según se manejen tasas o densidades de incidencia, respectivamente. Si la incidencia se mide como densidad, el RA se divide por unidades de tiempo. En el caso del tabaco y el cáncer de pulmón (Tabla 2) un RA = 195/100.000 significa que el tabaco produce casi dos muertes por cáncer de pulmón al año entre cada 1.000 fumadores; esta es la mortalidad por cáncer de pulmón que se evitaría erradicando el tabaco en este colectivo. Una medida derivada del RA es la proporción de riesgo atribuible (RA %), también llamada fracción etiológica. Es el RA dividido por la incidencia de enfermedad en los expuestos (Tabla 1). Es otra forma más de presentar el impacto del factor de riesgo entre los expuestos a él. Expresado en términos útiles para la prevención, representa la proporción de la incidencia de enfermedad que se evitaría entre los expuestos si se erradicara el factor de riesgo. Al ser una proporción carece de unidades, su límite inferior es 0 y el superior es 1 (100 si se expresa como porcentaje). En nuestro ejemplo (Tabla 2), un RA % = 93,3 % significa que el 93,3 % de los cánceres desarrollados por los fumadores durante los 40 años del seguimiento se deben al consumo de tabaco. Por lo tanto, ese mismo porcentaje se habría evitado si no hubieran fumado. También puede estimarse el número de sujetos cuyo hábito tabáquico habría que interrumpir para evitar una muerte o evento adverso; es el llamado índice de intervención (II).  $II = 1/RA$ . En el ejemplo, II es 512, y significa que hay que evitar el hábito tabáquico en 512 fumadores para evitar una muerte por cáncer de pulmón cada año. En el contexto clínico, el II también se conoce simplemente como el “número de sujetos a tratar” para evitar un evento indeseable.

Si se trata de un factor protector, puede calcularse una medida análoga a la RA %, la llamada fracción preventiva (FE), que representa la proporción de la incidencia de enfermedad evitada entre los expuestos por la acción del factor protector:

$$FE = (I_0 - I_e) / I_0 \times 100 = (1 - RR) \times 100$$

donde "I<sub>0</sub>" es la incidencia en los no expuestos, y "I<sub>e</sub>" la incidencia en los expuestos al factor de riesgo.

El riesgo atribuible en la población general (RAP) es la diferencia de la incidencia de la enfermedad entre la población general y los no expuestos al factor de riesgo. También cabe calcularlo multiplicando el RA por la prevalencia del factor de riesgo en la población general (Tabla 1). Tiene las mismas unidades y límites que el RA. Es una medida del impacto del factor de riesgo en la población general, que ha de ser menor que en los expuestos al factor de riesgo, al estar formada por expuestos y no expuestos a dicho factor. Representa el “exceso” de incidencia de enfermedad que se evitaría en la población general si se erradicara el factor de riesgo. Es, junto con la RAP %, la medida de efecto de mayor importancia en salud pública. En nuestro ejemplo (Tabla 2), un RAP de 127/100.000 mide la reducción de la mortalidad anual por cáncer de pulmón en el conjunto de la cohorte si se erradicara el consumo de tabaco. La proporción de riesgo atribuible en la población general (RAP %) es una medida derivada del RAP y se obtiene dividiendo este por la incidencia total de enfermedad en el conjunto de la población (Tabla 1). Representa la proporción de la incidencia de enfermedad que se evitaría en la población general si se erradicara la exposición al factor de riesgo. La RAP % carece de unidades, su límite inferior es 0 y el superior es 1 (100 si se expresa como porcentaje). En el caso del tabaco y el cáncer de pulmón, la RAP % = 90,1 % significa que la erradicación del consumo de tabaco reduciría la mortalidad por cáncer de pulmón en la cohorte de estudio en un 90,1 %. Es posible también calcular el número de muertes atribuibles al tabaco multiplicando la RAP % por el número de muertes por cáncer de pulmón.

Un factor de riesgo con un RR modesto puede tener un gran impacto sobre la salud pública, una elevada RAP %, si su prevalencia es muy alta. Es el caso de la tensión arterial y su impacto sobre la cardiopatía isquémica; su RR es aproximadamente 2, pero alrededor del 20 % de la población presenta este factor de riesgo. Por el contrario, el asbesto tiene un RR superior a 10 sobre la mortalidad por cáncer de pulmón, pero su impacto sobre la mortalidad por esta enfermedad es solamente moderado, porque su prevalencia en la población general es muy baja. Sin embargo, tiene un enorme impacto sobre los expuestos profesionalmente a este material.

#### ■ 6.1.4. Variabilidad en medidas de asociación

Las medidas de asociación descritas tienen una variabilidad derivada de la selección muestral en los estudios, y es posible calcular pruebas de significación estadística e intervalos de confianza para ellas. Por otro lado, el análisis de los datos debe iniciarse comprobando la comparabilidad de los grupos de expuestos y no expuestos, es decir, asegurando que son homogéneos en la distribución de otros predictores de la variable resultado, distintos del factor que se ha de evaluar. Si no es así, hay que ajustar las medidas de asociación por la acción de dichos predictores mediante métodos estadísticos, como la regresión multivariante (ver el apartado de dificultades en estudios observacionales y técnicas para solventarlos).



### ■ 6.1.5. Un ejemplo

A modo de ejemplo, la Tabla 4 presenta los resultados del efecto de una intervención en un ensayo clínico. En concreto se ha usado el ensayo SPRINT que se diseñó para establecer si, en pacientes hipertensos de alto riesgo cardiovascular, reducir la tensión arterial sistólica por debajo de 120 mm Hg era beneficioso por comparación con llevarla simplemente por debajo de 140 mm Hg (el valor habitual por debajo del que se considera que la hipertensión está controlada) (Doll *et al.*, 1994). Se reclutó a 9.361 pacientes, de los que 4.678 fueron asignados aleatoriamente al grupo de intervención (tensión <120 mm Hg) y 4.683 al grupo de control (tensión <140 mm Hg). La duración del ensayo fue de 3,26 años, y en el grupo de intervención se recibieron como media tres fármacos antihipertensivos y en el de control solo dos fármacos. La Tabla 4 presenta los resultados para una de las variables de resultado del ensayo, en concreto la mortalidad general. El RR fue 0,73, que indica que la intervención redujo la mortalidad un 27 % respecto al control. El RA fue 1,2 %, que indica que la intervención redujo en términos absolutos la mortalidad un 1,2 % a lo largo de los 3,26 años de seguimiento; ello supuso que bajar la tensión arterial por debajo de 120 mm Hg evitó 55 muertes. Por último, el número de pacientes a los que hubo que reducir la tensión a <120 mm Hg durante 3,26 años para evitar una muerte fue de 83. En conjunto la intervención produjo unos beneficios sustanciales sobre la mortalidad.

Tabla 4.

#### MEDIDAS DE EFECTO EN EL ENSAYO CLÍNICO SPRINT SOBRE REDUCCIÓN DE LA PRESIÓN ARTERIAL EN HIPERTENSOS Y MORTALIDAD GENERAL A LO LARGO DE 3,26 AÑOS

Mortalidad en el grupo de intervención (tensión arterial <120 mmHg): 3,3 %

Mortalidad en el grupo de control (tensión arterial <140 mmHg): 4,5 %

#### Medidas de asociación:

**Riesgo relativo:**  $3,3 \% / 4,5 \% = 0,73$

**Riesgo atribuible:**  $3,3 \% - 4,5 \% = -1,2 \%$

**Número de pacientes que hay que tratar para evitar una muerte:**  $1 / 1,2 \% = 83$

Fuente: Estimado a partir de SPRINT Research Group *et al.* (2015).

## ■ 6.2. CONTRASTES DE HIPÓTESIS Y REGRESIÓN LINEAL

### ■ 6.2.1. Introducción

El contraste de hipótesis (CH) es la técnica estadística que se utiliza para decidir entre dos posibilidades. Ejemplos de uso podrían ser:

1. Tras un ensayo clínico, decidir si el resultado obtenido en el grupo que ha utilizado el tratamiento test es igual al resultado obtenido para el grupo que ha utilizado el tratamiento convencional o un placebo.
2. Se conoce el porcentaje de supervivencia a determinada enfermedad ( $p_0$ ). Se hace un estudio con un nuevo medicamento obteniendo un porcentaje de supervivencia ( $p_1$ ). ¿Es mejor el nuevo medicamento?

Los dos ejemplos anteriores, habituales en ensayos clínicos, muestran el problema principal del contraste de hipótesis. Supongamos en el ejemplo 2 que la probabilidad de supervivencia actual es del 30,00 % y con el nuevo medicamento hemos obtenido una supervivencia de 30,0000001 %. Casi no es necesario saber estadística para afirmar que no hay diferencias entre ambos grupos. Del mismo modo si hubiéramos obtenido un valor de supervivencia para el grupo test de 90 % decidiríamos que el tratamiento test es diferente del control y además es magnífico.

El problema surge, en la práctica, porque los resultados no son tan extremos y necesitamos herramientas que nos permitan decidir si el nuevo medicamento es efectivo o no. El CH nos va a permitir formular estas preguntas de modo sistemático y obtener una respuesta en términos de probabilidad. Para ello es preciso plantear dos hipótesis:

$H_0$  Hipótesis nula. Es la hipótesis que se desea probar con los datos.

$H_1$  Hipótesis alternativa. Es, como su nombre indica, la alternativa a  $H_0$ .

Para realizar un contraste de hipótesis es preciso seguir los siguientes pasos:

1. Saber qué tipo de contraste se necesita.
2. Plantear correctamente  $H_0$  y  $H_1$ .
3. Obtención del p-valor.
4. Decisión.

Las etapas 1 y 2 son las más complejas hoy día, ya que requieren conocer bien el funcionamiento del problema y plantear las hipótesis correctamente. Las etapas 3 y 4 antiguamente requerían cálculos farragosos y el uso de cuadros de probabilidades. Actualmente son inmediatas ya que disponemos de *software* que realiza los cálculos (etapa 3) y, una vez obtenido el p-valor, de criterios claros sobre cómo tomar las decisiones. Ya no es necesario conocer la formulación matemática del contraste.

La decisión se toma en función del p-valor que nos da información de cuán verosímil es la hipótesis nula con los datos obtenidos. Cuanto menor sea el p-valor, tendremos menos evidencia de que nuestros datos sean compatibles con  $H_0$ . El criterio más utilizado es que si el p valor es menor que  $\alpha=0,05$  (5 %) rechazamos la

hipótesis nula. Es decir, pensaremos que es cierta la hipótesis alternativa. Si, por el contrario, el p-valor es mayor que 0,05 no rechazaremos la hipótesis nula. Es importante resaltar que el p-valor *da información sobre lo verosímil que puede ser  $H_0$ . No es la probabilidad de  $H_0$* . Este tema se trata en el apartado 6.2.3.

A continuación, se van a plantear cuatro tipos de contraste y su resolución con un ejemplo en cada caso.

## ■ 6.2.2. Tipos de contraste y cómo se plantean $H_0$ y $H_1$

### ■ 6.2.2.1. Contraste para una proporción

El ejemplo tipo de este contraste sería:

*Se sabe que la supervivencia a determinada enfermedad es de 0,3 (30 %). Se prueba una nueva medicación en una muestra  $n=56$  pacientes. Sobreviven a la enfermedad 19 (probabilidad de sobrevivir  $19/56=0,3393$ ). La pregunta claramente es: ¿Es mayor la tasa de supervivencia con la nueva medicación?*

La hipótesis nula siempre es la información externa que se quiere probar. En este caso,

$$H_0 : p=0,3$$

La hipótesis alternativa es igual a la nula siempre, salvo por el signo. Y el signo se decide comparando los datos muestrales con la hipótesis nula. Por tanto,

$$H_1 : p = 0,3$$

En nuestro caso hemos obtenido una proporción mayor de supervivientes (0,3393), por tanto, la alternativa es que la probabilidad de sobrevivir es mayor que 0,3,

$$H_1 : p > 0,3$$

El contraste se plantea completo de la siguiente manera:

Contraste para una proporción, con

$$H_0 : p=0,3$$

$$H_1 : p > 0,3$$

Una vez planteado el contraste cualquier *software* estadístico nos va a pedir la  $p$  de  $H_0$  ( $p=0,3$ ), el signo de  $H_1$  ( $>$ ), el tamaño muestral ( $n=56$ ), el número de super-

vivientes (19) o la probabilidad de supervivencia ( $19/56=0,3393$  o porcentualmente 33,93 %) y el nivel de significación  $\alpha$ , que suele ser 0,05. El resultado es el p-valor.

La web Mathcracker<sup>1</sup> permite hacer los cálculos para obtener el p-valor. Introduciendo los datos requeridos, obtenemos p-valor=0,2606 en el apartado 4 de respuestas.

Como el p valor es superior a 0,05 no podemos rechazar la hipótesis nula y por tanto seguimos pensando que  $p=0,3$  con la nueva medicación. Es decir, no ha supuesto una mejora en la tasa de supervivencia.

### ■ 6.2.2.2. Contraste para una media

Un ejemplo de este contraste sería:

*Se sabe que la duración media de una enfermedad es 15 días. Se prueba un nuevo tratamiento en  $n=48$  pacientes. Se obtiene una duración media de 13,2 días y una desviación típica de 2,2 días. La pregunta es ¿Ha disminuido la duración de la enfermedad?*

La hipótesis nula es la información externa que se quiere probar. En este caso que la media ( $\mu$ ) es 15 días:

$$H_0 : \mu=15$$

La hipótesis alternativa, como se ha dicho en el apartado anterior, es igual a la nula siempre salvo por el signo, que se decide comparando los datos muestrales con la hipótesis nula. En nuestro caso la media de la muestra es de 13,2 días. Menor que 15. Por tanto:

$$H_1 : \mu<15$$

El contraste completo sería:

Contraste para una media con

$$H_0 : \mu=15$$

$$H_1 : \mu<15$$

El *software* estadístico nos va a pedir la  $\mu$  de  $H_0$  ( $\mu =15$ ), el signo de  $H_1$  ( $<$ ), el tamaño muestral ( $n=48$ ), la media obtenida (13,2) y la desviación típica obtenida (2,2). El resultado es el p-valor.

La web Mathcracker<sup>2</sup> permite hacer los cálculos para obtener el p-valor. Introduciendo los datos requeridos, obtenemos un p-valor=0,0016 en el apartado 4 de los resultados.

<sup>1</sup> <https://mathcracker.com/z-test-for-one-proportion.php>

<sup>2</sup> <https://mathcracker.com/t-test-for-one-mean.php>

Como el p valor es inferior a 0,05 rechazamos la hipótesis nula y por tanto hay evidencia estadísticamente significativa de que la duración de la enfermedad se reduce con el nuevo medicamento.

### ■ 6.2.2.3. Contrastes para comparar dos medias

El contraste para comparar dos medias resuelve el tipo de problema siguiente:

*Se realiza un ensayo clínico de un nuevo medicamento para combatir la epilepsia. El grupo de control está formado por 235 pacientes que toman la medicina tradicional. Han tenido en 180 días un promedio de 5,3 crisis con una desviación típica de 2,3. El grupo de test está formado por 243 pacientes. Han tenido una media de 5,8 crisis con desviación típica 2,5. La pregunta es ¿Son iguales ambos tratamientos?*

En el contraste de diferencia de medias la hipótesis nula es, casi siempre, que ambas medias son iguales. La alternativa es que no son iguales:

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

Los datos necesarios, que nos pedirá el *software*, serán, el signo de  $H_0$  (=), el de  $H_1$  ( $\neq$ ), los tamaños muestrales ( $n_1=235$ ,  $n_2=243$ ) las medias (5,3 y 5,8) y las desviaciones típicas (2,3 y 2,5).

La web Mathcracker<sup>3</sup> pide los datos anteriores y se obtiene un p-valor de 0,1065. No podemos rechazar la hipótesis nula y, ambos medicamentos son equivalentes.

### ■ 6.2.2.4. Contraste para comparar dos proporciones

El ejemplo sería:

*Se realiza un ensayo clínico de un nuevo medicamento para combatir la epilepsia. El grupo de control está formado por 235 pacientes que toman la medicina tradicional. En 180 días han tenido crisis 161 pacientes (68,51 %). El grupo de test está formado por 243 pacientes. Han tenido crisis 180 pacientes (74,07 %). La pregunta es ¿Son iguales ambos tratamientos?*

En el contraste de diferencia de proporciones la hipótesis nula es, casi siempre, que ambas proporciones son iguales. La alternativa es que no son iguales:

$$H_0 : p_1 = p_2$$

$$H_1 : p_1 \neq p_2$$

<sup>3</sup> <https://mathcracker.com/t-test-for-two-means.php>

Los datos necesarios, que nos pedirá el *software*, serán: el signo de  $H_0$  (=), el de  $H_1$  ( $\neq$ ), los tamaños muestrales ( $n_1=235$ ,  $n_2=243$ ) y el número de pacientes con crisis o sus proporciones.

La web Mathcracker<sup>4</sup> pide los datos anteriores y se obtiene un p-valor de 0,1787 en el apartado 4 de resultados. No podemos rechazar la hipótesis nula y ambos medicamentos son análogos.

### ■ 6.2.2.5. Otros contrastes

Existen muchos contrastes, pero la forma de trabajar con ellos es siempre la misma. Plantear las hipótesis nula y alternativa correctamente. Mediante un *software* adecuado obtener el p-valor y con él decidir qué hipótesis debemos rechazar.

Un contraste muy habitual es el ANOVA, Análisis de la varianza. En este caso se tienen datos de varios grupos y se quiere saber si los grupos son iguales o si al menos uno de ellos difiere del resto. Suele utilizarse en ensayos en los que se miden los efectos de varias dosis diferentes a varios grupos de pacientes. En este caso  $H_0$  será que todas las medias son iguales y  $H_1$  que al menos una de las medias es diferente. El p-valor se analiza exactamente igual que hasta ahora.

También es importante mencionar que hay muchos contrastes diferentes para resolver el mismo problema. Por ejemplo, para estudiar una diferencia de medias se puede utilizar también el contraste no paramétrico de Mann-Whitney y el p-valor se interpretaría igual.

### ■ 6.2.3. El mal uso del p-valor

Como se ha explicado, el p-valor nos da información de cuán verosímil es  $H_0$  con los datos que tenemos, pero no es la probabilidad de que  $H_0$  sea cierta.

Técnicamente el p-valor es la probabilidad de que si  $H_0$  es cierta, se obtenga una muestra más desfavorable a  $H_0$  que la que hemos obtenido.

En la práctica esto equivale a que podemos equivocarnos un 5 % de las veces (cuando  $\alpha=0,05$ ) en el sentido de rechazar  $H_0$  y pensar que hay diferencias entre los dos grupos. Es decir, si repetimos el mismo ensayo sin diferencias 100 veces, en cinco ocasiones en promedio obtendremos p-valores menores que 0,05 aunque no haya diferencias entre los grupos y, consecuentemente rechazaremos  $H_0$  y crearemos que hay diferencias entre los grupos. Por ejemplo, supongamos que realizamos el siguiente experimento absurdo:

*Se escogen 50 personas sanas de 20 años al azar. Se les divide en dos grupos aleatoriamente y se les mide la tensión. Al grupo 1 se le ha medido la tensión sentado en una silla azul. Al segundo en una silla verde. Si*

<sup>4</sup> <https://mathcracker.com/z-test-for-two-proportions.php>

*nos planteamos si hay diferencias en la tensión media de ambos grupos, obviamente esperamos un p-valor mayor que 0,05 ya que el color de la silla no debería influir en la tensión de esas personas. Ahora bien, si repetimos ese experimento 100 veces, habrá aproximadamente 5 en las que el p-valor saldrá menor que 0,05. Y la conclusión, en esos casos, será que el color de la silla influye.*

Eso no es un problema grave si hacemos un ensayo y medimos una variable, pues la probabilidad de detectar algo inexistente es del 5 % que es baja. Pero si hacemos un ensayo y con la misma muestra medimos 200 variables –por ejemplo, todo tipo de marcadores– y hacemos 200 contrastes de diferencia de medias entre el grupo de control y el de test, vamos a tener en promedio 10 test con p-valores menores que 0,05 que no son reales. Es decir, son falsos positivos. Y pensaremos que ese medicamento es efectivo para esas diez cosas, y ese es el problema del mal uso del p-valor (Baker, 2016). Hay dos formas claras de evitarlo: no utilizar la misma muestra para hacer muchos contrastes de hipótesis o bajar notablemente el nivel de significación  $\alpha$ . El problema en este segundo caso es que, si bien bajaríamos el número de falsos positivos, aumentaría el de falsos negativos y no se detectarían bien tratamientos que si son efectivos. Consecuentemente la mejor forma de evitar problemas es utilizar cada ensayo para el tema elegido y no como cajón de sastre para probar muchas otras posibilidades (Doll *et al.*, 1994).

#### ■ 6.2.4. Regresión lineal

La regresión lineal se utiliza para estudiar cómo depende una variable Y de otra variable X.



Ejemplos de uso serían:

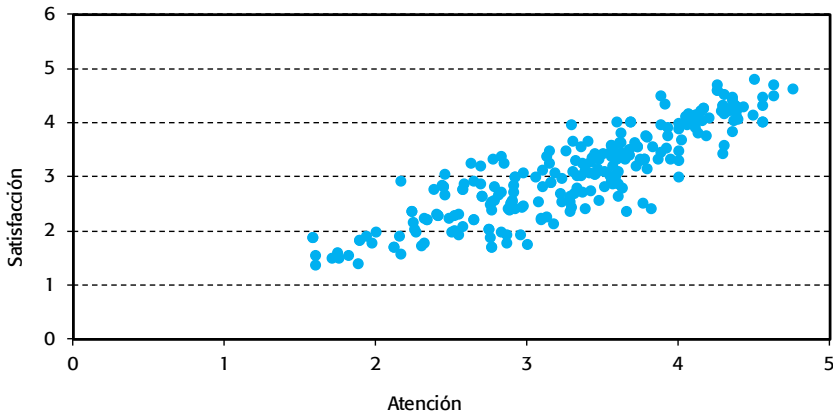
1. Calcular la concentración plasmática esperada (Y) de un medicamento en función de la dosis que se ha suministrado al paciente (X). Se tiene una muestra de pacientes y, para cada uno de ellos, se conoce la dosis recibida y la concentración plasmática.
2. Se realiza una encuesta de satisfacción de pacientes de un hospital. Se conocen, para cada paciente, los valores otorgados a la Satisfacción Global (Y) y a la Atención recibida (X).

El objetivo del análisis es estudiar dos aspectos:

- ¿Existe relación entre ambas variables?
- Cuantificar esa relación, es decir si se incrementa la dosis del medicamento en una unidad, ¿Qué incremento de la concentración plasmática cabe esperar?

Figura 1.

### DISPERSIÓN DE LA SATISFACCIÓN VERSUS ATENCIÓN RECIBIDA POR EL PACIENTE



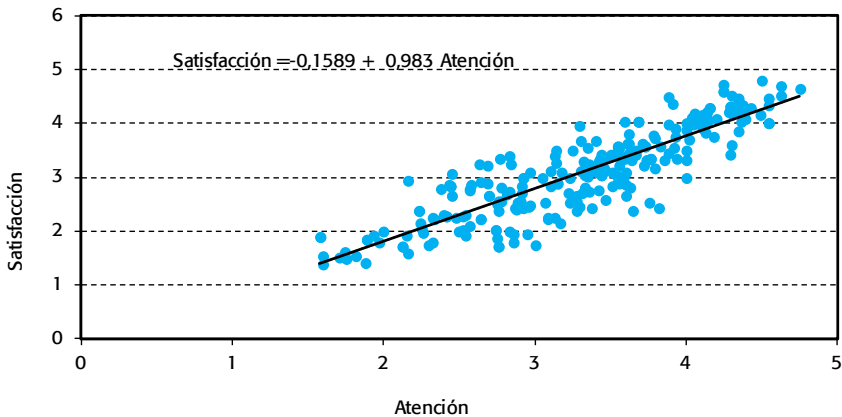
Fuente: Elaboración propia.

Se va a ilustrar el análisis con una muestra de datos del ejemplo 2. La Figura 1 muestra el gráfico de dispersión para los datos de satisfacción de los pacientes.

Como puede observarse en la figura, el gráfico presenta dos características importantes. En primer lugar, hay relación entre las dos variables, y en segundo

Figura 2.

### REGRESIÓN DE SATISFACCIÓN DE LOS PACIENTES EXPLICADA POR LA ATENCIÓN RECIBIDA



Fuente: Elaboración propia.



lugar esta relación es de tipo lineal. En estas circunstancias es razonable ajustar una línea recta a los datos y utilizar su ecuación como resumen de la relación entre ambas variables. Esto es lo que hace una regresión simple o regresión lineal. Si la ajustamos a los datos de la Figura 1, obtenemos la regresión de la Figura 2.

La ecuación matemática de la recta que aparece en la Figura 2 es:

$$\text{Satisfacción} = -0,1589 + 0,983 \text{ Atención}$$

Como ya indicamos al hablar de contrastes de hipótesis, hoy día no es necesario conocer las fórmulas que permiten obtener este resultado, ya que esta ecuación puede obtenerse con cualquier *software* estadístico e incluso con Excel.

En general la recta de regresión se escribe:

$$Y = \beta_0 + \beta_1 X$$

y, como puede observarse, en nuestro caso  $\beta_1=0,983$ . Esto indica que si el valor de X se incrementa en una unidad, el valor de Y se incrementará en 0,983 unidades. Es decir, si un paciente da un punto más en la encuesta a Atención recibida en el Hospital su Satisfacción Global se incrementará en 0,983 puntos.

Si los datos fueran del ejemplo de las dosis, la conclusión hubiera sido que si un paciente recibe una dosis una unidad mayor, su concentración plasmática habría aumentado en 0,983 unidades.

Es muy importante resaltar que la regresión nos permite cuantificar el impacto que los cambios en el valor de X tendrán sobre el valor de Y.

#### ■ 6.2.4.1. Contraste de significatividad de $\beta_1$

Cuando se ajusta una regresión es importante realizar el contraste de significatividad sobre el coeficiente  $\beta_1$ . Este contraste sería:

$$H_0 : \beta_1=0$$

$$H_1 : \beta_1 \neq 0$$

Si  $\beta_1=0$ , querría decir que la ecuación quedaría como:

$$Y=\beta_0+0 X=\beta_0.$$

Es decir que Y no depende de X. O dicho al revés, que X no influye sobre Y.



El contraste que resuelve esta cuestión se denomina Contraste t. Los programas que realizan regresiones proporcionan el p-valor del contraste t. Si el p-valor es menor que 0,05, concluiremos que X sí influye sobre Y y procederemos a cuantificar su efecto. En este caso hablamos de una variable *estadísticamente significativa*. Si, por el contrario, el p-valor es superior a 0,05 la variable no es significativa y podemos afirmar que X no influye sobre Y.

En nuestro ejemplo el p-valor=0,000013 que es menor que 0,05. Por ello concluimos que la variable X es significativa y sus cambios repercuten sobre Y.

#### ■ 6.2.4.2. Coeficiente de Determinación $R^2$

Finalmente, en regresión se obtiene el valor  $R^2$ , que indica cuánta variabilidad de Y es explicada por X. En el ejemplo que estamos utilizando  $R^2=0,89$ , es decir, 89 %. Podemos afirmar por tanto que X explica el 89 % de la variabilidad de Y. El valor de  $R^2$  depende básicamente del tipo de fenómeno que se esté estudiando, ya que hay situaciones en las que un 20 % de capacidad explicativa es mucho y otras en que un 60 % es poco. El analista normalmente conoce los rangos de capacidad explicativa esperables.

#### ■ 6.2.4.3. Otros modelos de regresión

Existen otros modelos de regresión que no se van a analizar en este capítulo como:

- La regresión lineal múltiple que es similar a la regresión lineal simple (visto en los apartados anteriores, con una variable Y y una variable X) pero en la que el modelo de regresión tiene más de una variable X, en ellas se analizaría el efecto de cada variable X sobre la variable Y controlado por las otras variables Xs.
- Regresión logística que se utiliza cuando la variable Y es una variable dicotómica cualitativa o categórica.
- Regresión de Cox que es un modelo utilizado cuando la variable de respuesta Y representa duraciones de procesos (por ejemplo, de enfermedades o tiempo de supervivencia de pacientes a una enfermedad mortal) y se desea conocer qué variables influyen en esa duración. El modelo se utiliza cuando parte de los datos de tiempos no se observan completos. Por ejemplo, porque un paciente sigue vivo en el momento del estudio. En ese caso, su tiempo de supervivencia real a la enfermedad será mayor que el que ha tenido hasta el momento de la toma de datos. Este tipo de datos se denominan censurados y el modelo de tasa de fallos proporcional de Cox es capaz

de trabajar con datos censurados proporcionando estimaciones centradas de los tiempos de supervivencia.

### ■ 6.3. LIMITACIONES DE ESTUDIOS OBSERVACIONALES Y ESTRATEGIAS DE ANÁLISIS

#### ■ 6.3.1. Necesidad de los estudios observacionales

Los estudios observacionales, basados en la obtención de datos de la práctica clínica real con pacientes, son necesarios, en especial en la evaluación de resultados en salud, ya que proporcionan información que no se puede obtener de los ensayos clínicos. Informan sobre los resultados en la práctica clínica fuera del contexto ideal, restrictivo y controlado, de los ensayos clínicos, pueden aportar información respecto a efectos raros o a largo plazo, sobre factores que afectan al uso o efecto de los tratamientos, efectos en un mayor grupo de población o con unas características diferentes a las incluidas en los ensayos, etc.

La realización de estudios observacionales es una labor que se debe realizar de forma paralela a la práctica asistencial para así contribuir a un mejor conocimiento de los efectos de los tratamientos, etc. Los profesionales sanitarios podemos contribuir en gran medida a ampliar y aplicar este conocimiento con orientación a la mejora de la práctica asistencial y de los resultados en los pacientes. Además, para el análisis de los registros clínicos, se realizan, en general, estudios observacionales, en muchas ocasiones de cohortes, que están ganando importancia para monitorizar y mejorar la calidad del cuidado de los pacientes (Gil Martin-Thomas, Liew y Hopper, 2018).

#### ■ 6.3.2. Principales limitaciones de los estudios observacionales

Sin embargo, los estudios observacionales tienen también limitaciones importantes que es necesario solventar para poder extraer conclusiones válidas. Entre estas limitaciones o dificultades se encuentran:

- La asignación no aleatoria de los individuos a las alternativas o factores de riesgo. En la práctica clínica es normal que los individuos hayan recibido una alternativa en lugar de otra por algunas características, disminuyendo la comparabilidad de los grupos.
- La falta de información: el limitado conocimiento sobre cómo se ha decidido la asignación a los tratamientos, la medida incompleta de los factores de riesgo que influyen en el tratamiento y en el efecto, etc. Es más, el carácter retrospectivo de muchos de estos estudios dificulta la recogida de cierta información. Además, la falta de datos puede no estar equilibrada entre los grupos

comparados, en general se recoge de forma menos completa en los individuos no tratados y, por ello, asumir que una condición no está presente cuando no está anotada puede ser problemático (Glynn, 2017).

Estas limitaciones pueden suponer sesgos e impedir establecer una relación causal entre una exposición o un factor de riesgo, ej un tratamiento, y un determinado efecto. El principal reto con los estudios observacionales para determinar relaciones causales válidas es solucionar en la medida de lo posible los posibles sesgos.

### ■ 6.3.3. Sesgos en los estudios observacionales

Sesgo es cualquier factor que pueda llevar a que los resultados se desvíen sistemáticamente de la realidad. Entre los sesgos podemos destacar los siguientes:

- Sesgo de selección: los grupos a comparar no han sido seleccionados de forma adecuada, por ello hay que tratar de evitar que se seleccionen individuos con una relación entre el factor y el efecto especial. Lo mejor es aleatorizar lo máximo posible. Ya que en estudios observacionales no se puede aleatorizar los individuos expuestos y no expuestos al factor, al menos se puede hacer una selección aleatoria de la muestra.
- Sesgo de información: se produce cuando no se recoge la información de todos los individuos o pacientes por igual o no se recoge la información de forma adecuada. Una buena estrategia es hacer que la persona que genera la información o la recoge no pueda manipularla, por ejemplo, sea “ciego” al grupo al que pertenecen los individuos. También es importante estandarizar la forma de recogida de los datos y las definiciones de estos, antes de recogerlos.
- Sesgo de indicación: se produce cuando existe una tendencia a que los individuos expuestos al factor, por ejemplo, que toman el medicamento evaluado, tengan unas características particulares, por las que se les ha dado el tratamiento. Este sesgo se evita con la aleatorización, pero, como hemos comentado, no es posible en los estudios observacionales. Es difícil de solucionar si claramente los expuestos al factor son diferentes a los no expuestos, porque nunca se conseguirá saber, por ejemplo, si se encuentra una asociación si ésta se debe al factor analizado o a otro. Sin embargo, si hay más individuos con unas características en un grupo que en otro, pero existen individuos similares en ambos grupos se puede corregir por el posible efecto de este sesgo. Una estrategia consiste en el uso de lo que se llama el índice de propensión, *propensity score* (PS) en inglés. Es un índice que indica la propensión a estar expuesto al factor analizado, por ejemplo, a reci-

bir un tratamiento, en base a unas características de los individuos. Cada vez se utiliza más esta técnica y por ello la desarrollaremos más adelante.

Por otro lado, existen lo que se conoce como factores de confusión y es necesario corregir su efecto para llegar a conclusiones válidas. Los factores de confusión son factores relacionados con el factor de exposición analizado, por ejemplo, el tratamiento, y relacionados con el efecto estudiado, pero no son la vía por la cual el factor produce el efecto. Por ejemplo, si se estudia la relación entre el ejercicio y el fallo cardíaco, fumar podría ser un factor de confusión, porque las personas que hacen ejercicio, en general, fuman menos, y fumar aumenta el riesgo de fallo cardíaco, y se podría concluir que los que no hacen ejercicio tienen más riesgo de fallo cardíaco; pero también podría ser porque fuman más, por lo que se tendría que corregir por este factor, fumar o no. La mejor forma para controlarlos es la aleatorización. Sin embargo, en los estudios observacionales, que no se puede aleatorizar, hay que corregir por los factores de confusión, aunque no se podrá hacer por los que no se conozcan. Existen diferentes estrategias para corregir el efecto de los factores de confusión, tanto en el diseño de los estudios (aleatorización, exclusión de pacientes con ciertas características, emparejamiento) como en el análisis de los resultados (estratificación o análisis de subgrupos, realización de modelos), cada una de estas estrategias tiene sus ventajas y sus inconvenientes (Tabla 5).

Tabla 5.

**ESTRATEGIAS DE CONTROL DE LOS FACTORES DE CONFUSIÓN**

Estrategia	Ventajas	Inconvenientes
<i>En el diseño del estudio</i>		
Randomización	<ul style="list-style-type: none"> <li>• Garantiza la comparabilidad</li> <li>• Aumenta eficacia del estudio</li> </ul>	<ul style="list-style-type: none"> <li>• No es posible en estudios observacionales</li> </ul>
Emparejamiento	<ul style="list-style-type: none"> <li>• Aumenta eficiencia, precisión</li> </ul>	<ul style="list-style-type: none"> <li>• Alarga tiempo</li> <li>• Problemas factibilidad</li> </ul>
Restricción o exclusión	<ul style="list-style-type: none"> <li>• Simple</li> <li>• Aumenta homogeneidad</li> </ul>	<ul style="list-style-type: none"> <li>• Alarga tiempo</li> <li>• Reduce validez externa</li> </ul>
<i>En el análisis de los resultados</i>		
Estratificación	<ul style="list-style-type: none"> <li>• Simple</li> </ul>	<ul style="list-style-type: none"> <li>• Se puede perder información</li> <li>• Número limitado de categorías</li> <li>• Interpretación compleja</li> </ul>
Modelos	<ul style="list-style-type: none"> <li>• No se pierde información</li> <li>• No se categoriza</li> <li>• Puede dar estimaciones individuales de riesgo</li> </ul>	<ul style="list-style-type: none"> <li>• Se asumen condiciones no siempre válidas</li> <li>• Resultados difíciles de interpretar</li> </ul>

### ■ 6.3.4. Análisis aplicando el índice de propensión

Los pasos para la realización de análisis de los datos utilizando el índice de propensión son los siguientes:

1. Analizar las características de los pacientes en ambos grupos y ver cuáles son diferentes.
2. Crear un modelo que permita calcular el índice de propensión (PS) o la probabilidad o riesgo de que un individuo por sus características haya sido asignado a un grupo, por ejemplo, a recibir un fármaco u otro fármaco. Esto se realiza generalmente mediante un modelo de regresión logística ( $Y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$ ). En este modelo la variable dependiente (Y) será la asignación al grupo, por ej. a la intervención a evaluar sí=1 o no=0. Las variables independientes,  $x_1, x_2, x_3, \dots, x_n$ , serán variables que puedan estar relacionadas con la asignación a un grupo u otro y/o con el resultado que queramos medir. Como variables x se incluyen variables con sentido biológico, intentando no dejar fuera ninguna variable que pueda introducir un sesgo.
3. Una vez construido el modelo de regresión logística, calcular el valor del PS (Y) para cada individuo del estudio en función de los valores de las variables  $x_1, x_2, x_n$  de ese individuo en concreto (el PS será un valor entre 0 y 1). El PS nos indica la propensión de ese individuo, por sus características, a haber sido asignado al grupo de la alternativa/intervención evaluada.
4. Crear grupos comparables mediante emparejamiento utilizando los valores de PS o hacer modelos de regresión con el PS como una variable X. Al hacer emparejamiento, o grupos comparables en los valores de PS, se hacen los grupos comparables en las otras variables  $x_1, x_2, x_3, \dots, x_n$ . Para ello se pueden utilizar diferentes técnicas, las principales diferencias entre ellas están en el número de individuos que quedan tras el ajuste o emparejamiento y el peso relativo que se da a cada uno de los individuos. En ocasiones el PS se utiliza como una variable predictora o de confusión de un modelo (una variable independiente del modelo de análisis del efecto de la intervención en una determinada variable de resultado, ej  $Y = a + b_1x_1 + b_2x_2$ ; donde sería Y la variable de resultado y  $x_1$  la intervención y  $x_2$  el PS, y se podrían añadir también otras variables).

El emparejamiento más común consiste en emparejar cada individuo del grupo con el factor analizado, ej. el tratamiento estudiado, con el individuo del grupo control más similar en PS en una relación k:1. El emparejamiento puede ser exacto incluyendo el mismo número de individuos con cada valor del índice de propensión en cada grupo, o emparejar en función del PS con un límite de tolerancia fijado (caliper) o incluir todos los individuos del grupo control con diferentes pesos en función de su PS (densidad de kernel) y así se pierden menos pacientes, pero en estos casos también conviene limitar

el análisis a los individuos con valores comunes de PS. Otro aspecto para considerar es el reemplazamiento, es decir, si un mismo individuo del grupo control puede ser usado como control de más de un individuo del grupo de intervención. El emparejamiento puede ser total o por estratos o subgrupos (Stuart, 2010).

5. Si se utilizan técnicas de emparejamiento, es necesario comprobar que las muestras son comparables en cuanto a sus características tras el emparejamiento. Si no se consigue, se debe descartar ese método de emparejamiento y realizar otro nuevo hasta que las muestras estén balanceadas en las covariables. Algunos autores recomiendan estudiar la diferencia estandarizada de medias (diferencia de medias dividido por la desviación estándar global) que sea  $<0,25$  (Stuart, 2010), las gráficas de quintiles-quintiles y las probabilidades de las pruebas de comparación (ej. kolmogorv-Smirnov, bootstrap p value). Valorar si no va a ser posible separar las covariables de la asignación al tratamiento impidiendo el análisis.
6. Comparar el efecto en ambos grupos mediante las técnicas estadísticas convencionales. Se pueden usar modelos. Estimar la diferencia entre los tratamientos y su intervalo de confianza del 95 % (IC95 %). Para el cálculo de estos intervalos de confianza se pueden utilizar técnicas paramétricas, pero en muchas ocasiones es necesario recurrir a técnicas no paramétricas como el bootstrap (ver más abajo).

### ■ 6.3.5. Ventajas y desventajas del índice de propensión y otras estrategias

Las ventajas del uso de PS es que se simplifica la corrección del sesgo de indicación o la corrección del efecto de los factores de confusión. Por ejemplo, si se quiere hacer emparejamiento por todas las posibles variables de confusión serían muy complicado ya que habría que buscar en ambos grupos individuos con las mismas características. Sin embargo, hacer emparejamiento por el índice de propensión es más sencillo. Pero la desventaja es que solo permite corregir por las variables conocidas y medidas. Por lo tanto, es peor que la aleatorización.

Otras estrategias de corrección de estos posibles sesgos en los estudios observacionales son los métodos de *genetic matching* (GenMatch), que se puede realizar con el programa R. GenMatch es un algoritmo automático que maximiza el balance entre los grupos, busca el balance no sólo en el PS pero también en otras variables, y puede usar o no el PS. Requiere elegir las variables más importantes para buscar el balance entre los grupos, lo cual exige un criterio clínico y no solo estadístico, y el criterio para buscar la diferencia, para reemplazamiento, etcétera.

Tanto PS como GenMatch corrigen solo el efecto de las variables observadas, pero pueden existir variables no observadas que introduzcan un sesgo. En

ocasiones se realizan análisis de sensibilidad para analizar la posible confusión por factores no observados. Por ejemplo, analizar qué magnitud debería tener el efecto de las variables no observadas para cambiar las conclusiones y comparar esta magnitud con el efecto del factor de confusión observado más importante, si es menor asumir que es difícil que esto ocurra, es difícil que una variable no observada cambie las conclusiones y, por tanto, se estima que el resultado es robusto. Otros métodos incluyen técnicas de *machine learning*, corrección por variables instrumentales (variables relacionadas con la asignación al tratamiento, pero no con el resultado, por ejemplo, la distancia al hospital), etcétera.

### ■ 6.3.6. Bootstrap

Bootstrap es una técnica de remuestreo que genera múltiples muestras a partir de una muestra. Para generar cada una de estas muestras, se seleccionan aleatoriamente valores de la muestra de partida con posibilidad de repetirse algunos de ellos, ya que se generan muestras normalmente con el mismo número de individuos que la muestra inicial. En cada muestra que se va generando se toman individuos diferentes de forma aleatoria, pero todos incluidos en la muestra de partida. Mediante esta técnica se puede calcular la estimación del efecto (ej. diferencia entre grupos, *odds ratio* (OR), RR) en cada una de las muestras creadas y repetir el mismo proceso  $n$  veces, creando “ $n$ ” valores de estimación del efecto, que seguirán una distribución normal y para el que podremos calcular la media y su intervalo de confianza del 95 % (IC95 %). En el caso de estimación de una OR, por ejemplo, en un modelo de regresión logística  $Y = a + b_1x_1 + b_2x_2 + \dots$ . Se puede estimar mediante bootstrap el valor de  $b_1$ , y como  $e^{b_1}$  es la OR corregida por los otros factores ( $x_2 \dots$ ), podemos estimar la OR mediante bootstrap y su IC95 %. Bootstrap es especialmente interesante para el cálculo del IC95 %.

### ■ 6.4. CONCLUSIONES

- Generalmente, el efecto de un factor de riesgo sobre una enfermedad se evalúa en estudios epidemiológicos y el efecto de una intervención sanitaria sobre una enfermedad mediante ensayos clínicos.
- Los estudios observacionales, basados en la obtención de datos de la práctica clínica real son necesarios en la evaluación de resultados en salud.
- Las medidas de asociación entre el factor de riesgo/intervención sanitaria y la enfermedad; si la relación es causal, constituyen medidas del efecto del factor de riesgo/intervención sanitaria. Las más utilizadas son el riesgo relativo, la *odds ratio* y el riesgo atribuible.
- Las técnicas estadísticas, requeridas en cualquier investigación, permiten determinar si las relaciones causales o de asociación existen.



- Los contrastes de hipótesis son técnicas estadísticas que se utilizan para decidir entre dos posibilidades, por ej entre el grupo con y sin intervención, y obtener respuestas en términos de probabilidad.
- Los estudios no pueden analizar muchas variables, para evitar falsos positivos se debe utilizar cada estudio para analizar solo el tema para el que ha sido diseñado.
- Las principales limitaciones de los estudios observacionales son la asignación no aleatoria, disminuyendo la comparabilidad de los grupos, y la falta de información, pudiendo llevar asociados sesgos que impidan determinar una relación causal entre el factor y el efecto.
- Es necesario corregir el efecto de posibles factores de confusión para llegar a conclusiones válidas, para ello se pueden utilizar estrategias como la exclusión, el emparejamiento, la estratificación, los modelos, el índice de propensión.
- Cualquier estudio científico debe apoyarse en la investigación observacional o experimental de datos y utilizar las técnicas de análisis adecuadas a los mismos.
- Un investigador o analista debe tener muy claro las limitaciones de los estudios y utilizar estrategias para controlar su efecto.

## ■ REFERENCIAS

- BAKER, M. (2016). Statisticians issue warning over misuse of P values. *Nature*, 531(7593), p. 151.
- DOLL, R., PETO, R., WHEATLEY, K. *et al.* (1994). Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ*, 309(6959), pp. 901-911.
- GIL MARTIN-THOMAS, J. F. M., LIEW, D., HOPPER, I. (2018). Observational studies and their utility for practice. *Aust Prescr*, 41(3), pp. 82-85.
- GLYNN, R.J. (2017). Use of propensity scores to design observational comparative effectiveness studies. *J Natl Cancer Inst*, 109(8), p. djw345.
- SPRINT RESEARCH GROUP, WRIGHT, J, T. JR., WILLIAMSON, J. D., WHELTON, P. K. *et al.* (2015). A Randomized Trial of Intensive versus Standard Blood-Pressure Control. *N Engl J Med*, 373(22), pp. 2103-2116.
- STUART, E.A. (2010). Matching methods for causal inference: a review and look forward. *Stat Sci*, 25(1), pp. 1-21.