

INVESTIGACIÓN DE LA RELACIÓN CON O SIN CAUSALIDAD ENTRE VARIABLES

La mayor parte de los textos han sido tomados de:

Sofía B Cabezas Hernández, Manuel A Villa Vigil. Errores metodológicos clamorosos de las investigaciones en ciencias de la salud, biológicas y afines. Oviedo. KRK Ediciones. 2017. Anexo 2: Investigación de relaciones entre variables. p. 211-238.

I. INVESTIGACIÓN

Investigar es buscar un saber no conocido. La ciencia es un medio para investigar los hechos cuando éstos son de naturaleza experimental u observacional. Muy resumidamente, en ciencia hay un primer paso **ontológico** (descubrir el “ser” íntimo de los perceptos mediante el “des-cubrimiento” del mayor número posible de sus atributos), y un segundo paso **epistemológico** (“des-cubrir” las relaciones entre las piezas ontológicas anteriormente descubiertas).

1. El primer paso, ontológico, que también puede ser nombrado como descriptivo, pretende descubrir con la máxima precisión y exactitud posibles alguno o algunos de los atributos de los perceptos o fenómenos (objetos, individuos o poblaciones), como por ejemplo la frecuencia de un suceso.
2. El segundo paso, epistemológico, que también puede ser nombrado como relacional o analítico, busca las posibles relaciones existentes entre las diversas piezas ontológicas o variables, para: **1)** “explicar”, con el mejor modelo explicativo posible, las preguntas que surgen de los fenómenos observados; y **2)** posteriormente si cabe, intentar “predecir”, con el mejor modelo matemático posible, los valores desconocidos de unas, a través de los valores conocidos de otras.

INCONSCIENTES DE ESTAR ATRAPADOS POR UNA METONIMIA: A PROPÓSITO DE LA CIENCIA.

Suele incurrirse en un error ontológico, por abuso de la metonimia, al confundir “el ser de la ciencia” con “la actividad de los científicos”, en frases como “la ciencia ha conseguido bombardear al átomo y beneficiar o dañar a...”, cuando son los científicos en sus actividades los que consiguen eso. La ciencia es (quizás) la más noble de las construcciones humanas, pero su ámbito es únicamente intelectual, y no moral. En efecto, la ciencia no tiene ninguna intención, no toma decisiones, no prefiere hacer en lugar de no hacer, ni prefiere el mal sobre el bien ni viceversa, ni curar el cáncer o salvar vidas humanas a costa de vidas de otros animales, pues son los científicos los que tienen intención, prefieren, toman decisiones y ejercen acciones morales. Siendo una parte de la filosofía, la ciencia es el método intelectual de búsqueda de la verdad cuando las premisas son de naturaleza experimental u observacional. Cuando las premisas no son de esta naturaleza, para avanzar hacia la verdad sigue haciendo falta el resto de la filosofía.

En efecto, no puede ser la ciencia, sino el científico el agente moral que utiliza el conocimiento como un medio para alcanzar un fin (telos). Ese telos es un enunciado moral (prefiero hacer esto que eso) que constituye la premisa principal de un razonamiento práctico, tanto si el agente es consciente de ello como si no lo es. Los agentes que utilizan la ciencia como un medio, y que son conscientes del fin, expresan la premisa principal de facto (o serían capaces de expresarla) mediante una proposición del telos moral. Los demás agentes también persiguen sus fines, aunque se muevan en un nivel en el que no sean conscientes de que los persiguen, probablemente porque viven en la creencia (no reflexionada) de que “la ciencia” tiene un propósito, sin atisbar que los que tienen propósitos son los matemáticos criptógrafos, los físicos nucleares y los farmacólogos, y no la matemática ni la física ni la farmacología.

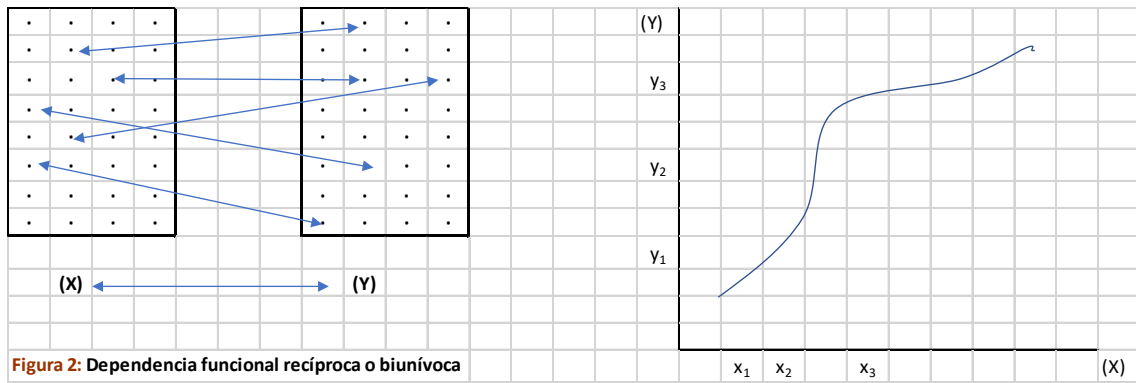


Figura 2: Dependencia funcional recíproca o biunívoca

2) Que a cada determinado valor de X le corresponda un único valor de la variable Y, pero no al contrario, porque a cada valor de Y le pueden corresponder dos o más valores de la variable X (o viceversa: a cada determinado valor de Y le corresponde sólo uno de X, pero a cada valor de X le corresponde más de uno de Y), tal como se muestra en la **figura 3**.

Esto significa una dependencia funcional unívoca, o implicación unívoca, $X \Rightarrow Y$.

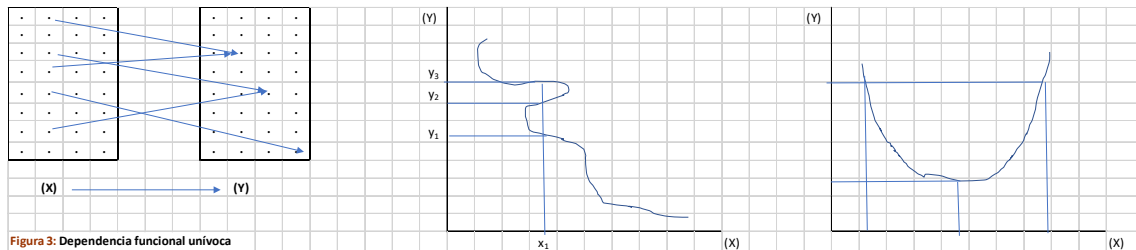


Figura 3: Dependencia funcional unívoca

2. De una manera flexible (situación aleatoria o estocástica), en la que a cada valor de una variable le pueden corresponder varios valores próximos de la otra, **debido a la influencia de otras variables ajenas, que englobamos bajo el nombre de azar, acaso, albur, chamba, casualidad**, etc. A su vez esta relación aleatoria o estocástica puede ser:

1) Relación aleatoria recíproca o biunívoca [$X \Leftrightarrow Y$] (**figura 4**).

En lógica formal, cuando X implica Y, e Y implica X, puede demostrarse que X es condición necesaria y suficiente de Y, e Y es condición necesaria y suficiente de X.

Un ejemplo es el peso (X) y la altura (Y) de las personas.

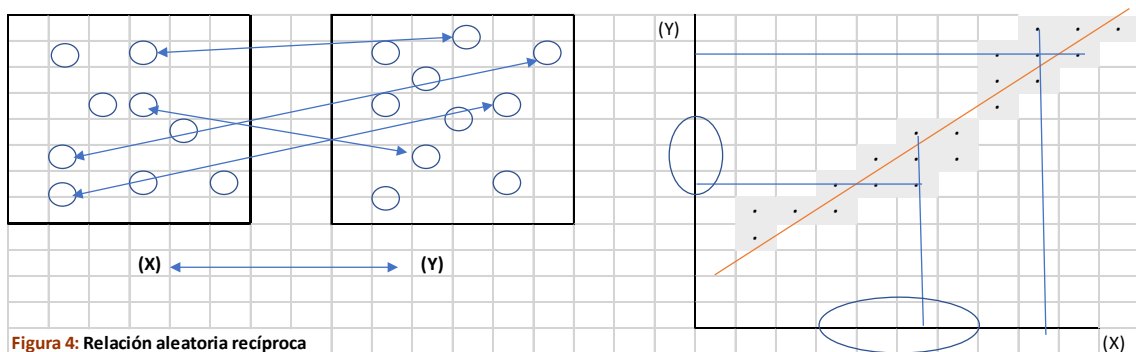
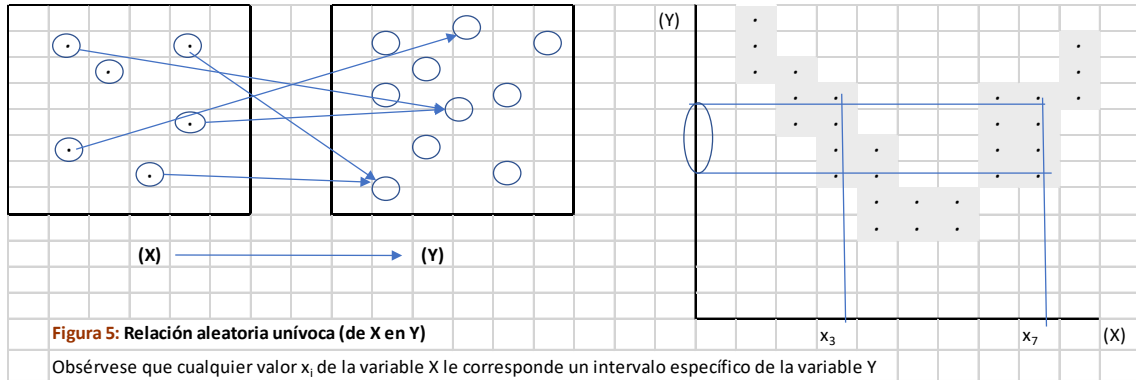


Figura 4: Relación aleatoria recíproca

2) Relación aleatoria unívoca del ser de X en el ser de Y [$X \Rightarrow Y$] (figura 5).

En lógica formal, cuando X implica Y, puede demostrarse que X es condición suficiente de Y, e Y es condición necesaria de X. Puede afirmarse que Y es condición necesaria pero no suficiente de X.

Un ejemplo es cualquier curva en "U" o en "J", como las cifras de colesterol en sangre en el eje de las X frente a las tasas de mortalidad en el eje de las Y.

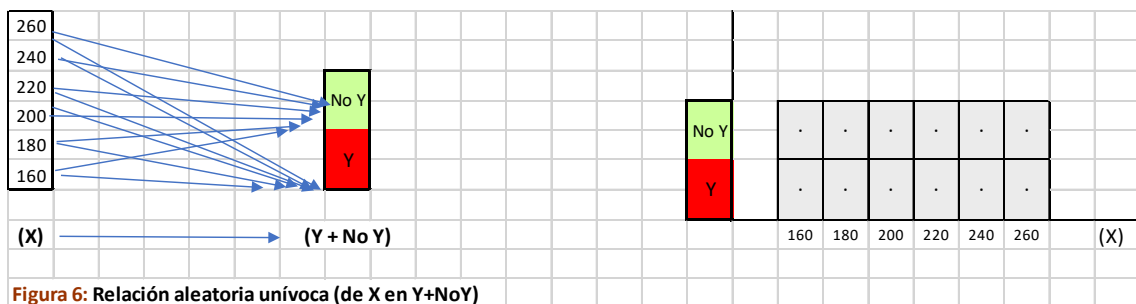


3) Relación aleatoria unívoca de "el ser de X en el ser de Y + el ser de no Y, [$X \Rightarrow Y+NoY$] (figura 6).

En lógica formal, cuando X implica Y+NoY, puede demostrarse que X es condición suficiente de Y + NoY.

Simultáneamente Y+NoY es condición necesaria de X, con lo que no puede afirmarse que Y sea condición necesaria de X, ni tampoco que NoY sea condición necesaria de X.

Un ejemplo es el de las cifras del colesterol en sangre que en un período de 10 años tienen los que mueren (Y) y los que no mueren (NoY).

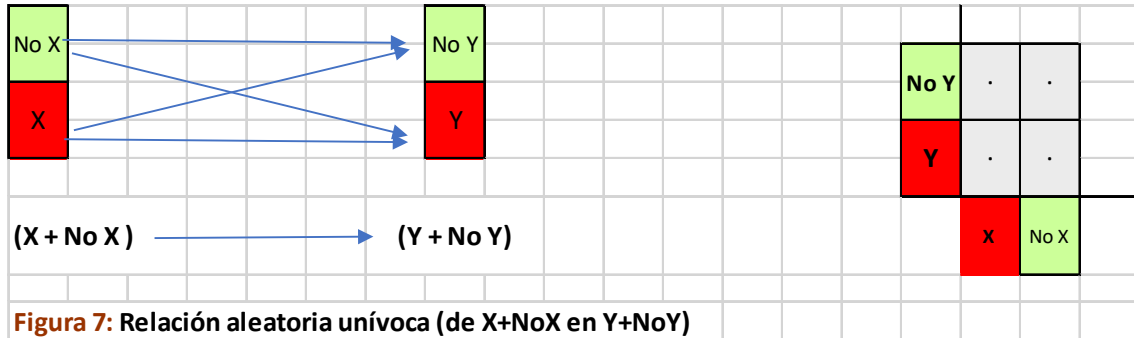


Cuando $Y+NoY = f(X)$ hemos de inferir que falta algo, y que más bien debe de tratarse de:
 $Y+NoY = f(X) +$ una o más causas independientes de X + las interacciones de estas u otras nuevas causas o factores con X y/o entre sí.

4) Relación aleatoria unívoca del ser de X + el ser de NoX en el ser de Y + el ser de no Y, $[X+NoX \Rightarrow Y+NoY]$ (figura 6).

X + NoX es condición suficiente de Y + NoY. Simultáneamente Y + No Y es condición necesaria de X + NoX, con lo que no puede afirmarse que Y sea condición necesaria de X ni de NoX, ni tampoco que NoY sea condición necesaria de X ni de NoX.

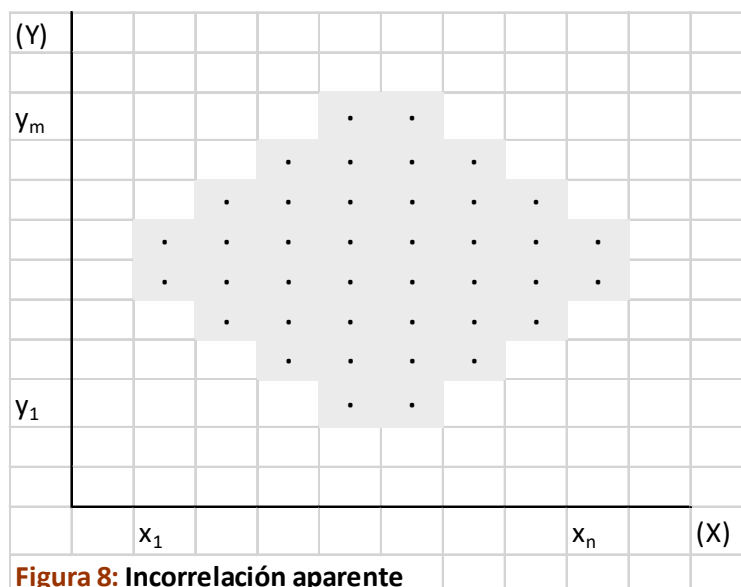
Un ejemplo es la asociación entre el colesterol alto (X) y colesterol no alto (NoX) en 10 años con los que mueren (Y) y los que no mueren (NoY). Véase tabla SCORE.



Cuando $Y+NoY = f(X+NoX)$ hemos de inferir que falta algo, y que más bien debe de tratarse de: $Y+NoY = f(X+NoX) +$ una o más causas independientes de X + las interacciones de éstas u otras nuevas causas o factores con X y/o entre sí.

2º Incorrelación aparente.

En la **figura 1** se ve la representación gráfica llamada diagrama de dispersión de dos variables incorreladas. Al compararlas con el diagrama de la **figura 8** ¿Están en él las variables de X e Y también incorreladas?



Desde luego, en el diagrama de dispersión de la figura 8 no se puede decir que las variables X e Y no estén relacionadas, porque:

1. El intervalo de los valores mostrados por la variable Y para el valor X_1 : **a)** no es el mismo que el intervalo de valores de la misma variable Y para el valor X_i de la variable X; y **b)** algo análogo cabe decir de los intervalos de valores de X para el valor de Y_1 y para el valor de Y_j .
2. Por lo que no se puede negar que hay alguna relación entre los valores de X y los valores de Y.

Sin embargo, también es verdad que la media de los valores de Y es la misma que los valores de X comprendidos entre X_1 y X_n (vemos que dibujan una línea recta horizontal, en la que el valor en Y no varía), y la media de los valores de X es la misma para todos los valores de Y comprendidos entre Y_1 e Y_m (vemos que dibujan una línea recta vertical, en la que el valor en la variable X es siempre el mismo), por lo que:

- 1) No son variables linealmente correlacionadas (como lo son en las **figuras 3, 4 y 5**, en las que se señalan las respectivas rectas de regresión lineal que resumen la correlación rectilínea entre las variables X e Y).
- 2) Pero sí que están relacionadas (que no es lo mismo que decir linealmente correlacionadas), ya que: **a)** los valores de X influyen en los valores de Y (mejor dicho, guardan relación con los valores de Y, porque decir que influyen es demasiado decir, habida cuenta de que no sabemos si son aspectos distintos cuyos valores casualmente se corresponden; y **b)** los valores de Y guardan relación con los valores de X.

Por lo tanto, son variables relacionadas, pero no correlacionadas (denominación que presupone que la relación lineal entre las variables X e Y, o sea, que se sigue la ecuación de una recta como $y = a \cdot x + b$, ó $x = r \cdot y + s$, en las que “a” y “r” son las constantes llamadas coeficientes o pendientes, y “b” y “s” son las constantes en el origen.

III. INVESTIGACIÓN DE RELACIONES ENTRE VARIABLES.

1º Entornos deterministas o no aleatorios.

Cuando dos variables se relacionan en un entorno determinista o no aleatorio (es decir, sin influencia de terceras variables), los valores de una se corresponden rígidamente con los de la otra. Por lo tanto, con buenas observaciones y buenas medidas (o sea, precisas, con poco intervalo de incertidumbre en los valores obtenidos en las medidas) y exactas (o sea, fidedignas), es fácil identificar las relaciones deterministas entre variables.

Los físicos, muy acostumbrados a entornos deterministas de la física clásica señalan que hay tres maneras de conseguir identificar la existencia de relaciones entre dos o tres variables:

1) Las tablas, en las que se pueden atisbar las relaciones mediante la inspección de los valores consignados en ellas.

2) Las gráficas (de las que hay muchos tipos, según para qué tipo de demostraciones) que hacen muy visibles e intuitivas esas relaciones.

3) Las leyes, que cuando son leyes matemáticas se traducen en ecuaciones, llamadas ecuaciones de regresión. Mediante ellas, a partir de los valores de una o varias variables, llamadas variables independientes, se puede deducir el valor de otra variable, conocida como independiente (a veces son varias, y entonces sus valores se agregan en forma de matriz).

Aunque las leyes pueden vincular a varias variables simultáneamente, las tablas y las gráficas sólo pueden tratar, de una manera ostensible, dos (a lo sumo, tres) de cada vez, presuponiendo que en ellas se mantienen constantes los valores de las demás variables estudiadas, porque si no, las tablas devienen en muy farragosas.

Las gráficas sólo permiten la inspección de tres dimensiones (en las representaciones llamadas espaciales, dibujadas en perspectiva), aunque son generalmente más claras con dos.

2º Entornos aleatorios o estocásticos.

El problema surge cuando se investiga la relación entre dos variables (en el caso más simple, aunque la situación puede complicarse cuando se estudian simultáneamente más) en un entorno aleatorio (que es lo habitual en ciencias de la salud), en el que el valor de una variable cuya dependencia de otras se quiere estudiar (por eso se llama variable dependiente): no sólo depende o se relaciona con los valores de las variables que estudiamos (variables dependientes), sino con otras variables desconocidas que influyen en los resultados de una medida que desconocemos.

A estas variables desconocidas “influyentes” se las llama variables de confusión y variables modificadoras del efecto o interacciones (aunque no son exactamente lo mismo, sus diferencias no son transcendentales para los objetivos de esta pieza pedagógica).

En unos casos, el conjunto de estas variables extrañas (o mejor dicho, desconocidas) tiene una repercusión contraria que en otras, debido a las diversas preponderancias posibles de unas o de otras, y del sentido de los efectos de cada una de ellas.

Precisamente por eso, las variables que sabemos relacionadas no exhiben un nexo o vínculo exacto y constante entre ellas (determinista), sino una “tendencia general”, o sea una relación aproximada, pero no perfecta, que dificulta la identificación de esas relaciones (vuelvan a verse las **figuras 4 y 5**).

A las situaciones que se dan en este entorno relacional se las llama aleatorias o estocásticas, porque parecen depender del azar, también llamado albur, acaso, casualidad, chanza, etc., que son los nombres que damos al conjunto de influencias o relaciones que

desconocemos o no controlamos, y que, por diversas proporciones posibles, unas veces repercuten de una manera y otras de otra.

3º Pruebas estadísticas en los entornos aleatorios o estocásticos.

Las relaciones aleatorias o estocásticas se estudian mediante pruebas estadísticas, o tests estadísticos, en los que se mide con algún indicador global cómo se distribuye o reparte el conjunto de valores correspondientes a cada una de las variables estudiadas en los distintos individuos componentes de una muestra representativa de la población que queremos investigar, para así ver cómo los valores de una variable se muestran ligados o vinculados con los de otra u otras.

Obsérvese que hemos dicho que esos valores de una variable “se muestran ligados o vinculados con los de otra u otras”, no que están causadas por ella/s, porque las relaciones entre los valores de dos variables X e Y, pueden ocurrir nexos que no son de naturaleza causal.

En resumen:

1. Puede haber una relación de causalidad, en la que:

1) los valores de X son causa de los de Y, en cuyo caso, X sería la variable independiente (pues sus valores son libres e independientes de los de Y), o variable causal (por cuanto que es la causa de los valores de Y); y

2) los valores de la variable Y son consecuencia de los valores de X (en cuyo caso, Y es la variable dependiente, a la que también podríamos llamar variable causada o efecto, por cuanto sus valores son causados por los valores de X); y

3) puede ser lo contrario, e incluso que sea difícil, o incluso imposible, determinar cuál es la causa y cuál es la consecuencia o efecto (por ejemplo, ¿quién es la causa de quién: el huevo o la gallina?).

2. También puede haber una relación de colateralidad, en la que las dos variables relacionadas, X e Y, no se causan una a otra, sino que están relacionadas por otra variable W, que puede ser, a su vez, causa (inmediata o directa, o mediata o indirecta) de las dos, y por eso sus valores “van del brazo” (de manera directa o indirecta).

3. Y también puede haber una relación de mera casualidad sin nexo de causalidad.

Dos sencillos ejemplos de la vida real ilustrarán ambas últimas circunstancias:

1) El profesor Maurice George Kendall encontró una estrechísima relación entre los incrementos anuales de licencias de radio concedidas entre 1924 y 1937 en EEUU, y el número de minusválidos psíquicos registrados entre el mismo período. Sin embargo, ambas son variables independientes entre sí, aunque están ligadas al crecimiento demográfico, y esa relación con éste les confiere una aparente relación entre ellas (colateralidad). El coeficiente de correlación de Pearson entre ambas variables fue 0,998, que es casi una correlación lineal determinista. Otros ejemplos como éste hemos visto en la web *spurious correlations*¹, que analizadas mediante el Coeficiente de Correlación Intraclase rebajan o pierden la correlación.

2) Entre 1965 y 1987 se observó en Alemania una poderosa relación entre la disminución de la natalidad y el retroceso en el número de parejas de cigüeñas. Evidentemente, fue una relación casual, sin ningún atisbo racional de causalidad.

Veamos un intuitivo ejemplo sobre cómo operan las pruebas estadísticas. Supongamos que tenemos dos bolsas o sacas opacas, a las que llamamos saca A y saca B, llenas de bolas blancas y negras. No sabemos cuántas hay de un color y del otro, ni el total, en cada saco. Simplemente sabemos, por su ruido al agitarlas que contienen bolas, y que, por su peso y su volumen, contienen muchas bolas, por lo que no es fácil ponerse a contarlas todas. Así que

¹ <http://www.tylervigen.com/spurious-correlations>

decidimos hacer una estimación mediante la extracción de un puñado de bolas (por ejemplo, de cinco, para hacer el problema más fácil) de cada bolsa como muestra.

Imaginemos que encontramos los siguientes resultados:

Bolsa A: 3 bolas blancas (60%) y 2 negras (40%)

Bolsa B: 2 bolas blancas (40%) y 3 negras (60%)

A juzgar por esos resultados, **parece** que hay más bolas blancas en A que en B (o más bolas negras en B que en A). Pero ¿es verdad o el hallazgo ha sido fruto de la casualidad en las extracciones de la muestra?

Supongamos que en la bolsa A realmente hubiera un 40% de bolas blancas y un 60% de bolas negras (o sea, lo contrario que lo que aparece en la muestra). ¿Es muy improbable que en la muestra saliera ese 60% de blancas y 40% de negras? Habría que calcular la probabilidad de extraer 3 bolas blancas y 2 negras de una saca que contiene el 40% de blancas y el 60% de negras. Y para esto hay fáciles procedimientos de cálculo: las pruebas o test estadísticos adecuados al tipo de variables y a la escala de medición (que no cualquiera, para no incurrir en un error tipo III).

Pasemos ahora a la bolsa B. Supongamos que en esa bolsa B hubiera el mismo número de bolas blancas que negras (50% de cada una). Evidentemente, al sacar cinco bolas, no es posible estimar ese 50% a 50%, porque implicaría sacar 2,5 bolas blancas y 2,5 bolas negras, lo que es imposible, porque las bolas con piezas indivisibles. O sea, que 3 bolas negras y 2 blancas, lo mismo que 3 blancas y 2 negras, serían las mejores estimaciones muestrales de la realidad poblacional, lo cual, si el experimento hubiera sido real no podríamos saberlo sin contar todas las bolas.

Podría decirse que hay que sacar 6 bolas, pero el argumento es erróneo por lo siguiente:

1) En nuestro ejemplo, hemos imaginado que hay 50% de blancas y 50% de negras, que no son porcentajes reproducibles en una muestra de 5 bolas, sino de 6 (en el supuesto de que el azar nos hubiera hecho sacar 3 bolas blancas y 3 bolas negras, lo que no es obligatorio) o de cualquier otro número par. En efecto podrían salir 0-6, 1-5, 2-4, 3-3, 4-2, 5-1 y 6-0.

2) Pero imaginemos que en la saca B (que es la cuestión en liza y no lo sabemos) hay un 52% de bolas blancas y el 48% de bolas negras. En ese caso, para reproducir en la muestra la composición real de la población (la saca), dicha muestra debería contar de 50 bolas como mínimo (o un múltiplo de 50), y además haber tenido la chiripa de que hubieran salido 26 blancas y 24 negras, pues ninguna otra posible combinación, salvo sus múltiplos, daría el resultado correcto, sino sólo un resultado aproximado.

Con 6 bolas, la mejor reproducción del contenido de la bolsa habría sido obtener 3 blancas y 3 negras, pues otras opciones serían estimaciones peores de la realidad.

Incluso si hubiéramos tenido la intuición de hacer muestras de 50 bolas, tampoco sabríamos si lo que realmente hay en la saca B:

1) No es, en vez de una proporción de 52% y 48%, un 50,5% y 49,5% (en cuyo caso sólo se obtendría con muestras de 200 bolas, o múltiplos de 200 bolas, si además la buena suerte nos proporcionara 101 blancas y 99 negras).

2) Ni tampoco sabemos si hay otra combinación de proporciones de bolas entre las infinitas posibles.

3) o incluso alguna presencia minoritaria de otro color diferente.

Puesto que las opciones son ilimitadas y desconocidas, no podemos saber a priori qué tamaño tiene que tener la muestra para poder reproducir exactamente la misma proporción que en la población.

Hasta podría ser que ningún tamaño muestral pudiera representar fidedignamente la composición de la población, como por ejemplo:

1) Que el tamaño de la población fuera un número primo, con lo que no podría ser dividido por ninguno de los números que expresan las respectivas frecuencias absolutas simples de bolas blancas y negras.

2) O que, sin ser primo, simplemente no fuera divisible por ninguna de dichas frecuencias absolutas simples de las bolas de cada uno de los colores, o sea, que fueran lo que en las matemáticas se llaman números primos entre sí (por ejemplo, 6 y 35 son primos entre sí, porque no tienen un número común por el que ambos sean divisibles, pero 6 y 27 no lo son porque ambos son divisibles por 3).

Así que habrá incertidumbre sobre la exactitud del valor de las muestras, con respecto a la población de origen, aun cuando el azar no nos hubiera dado lugar a una extracción de nuestra muestra desviada de la realidad.

Por todo lo dicho, en el ejemplo con las dos sacas de bolas, hemos vuelto al punto de partida, pues no sabemos:

1) Si en la saca A hay mayor proporción de bolas blancas que en la saca B, como se ha encontrado en la muestra.

2) Si en la saca A hay menor proporción de bolas blancas que en la saca B (o sea, al revés de lo que casualmente ocurrió en la muestra).

3) O si en la saca A hay la misma proporción de bolas blancas que en la saca B (lo que equivale a decir que en la saca A hay la misma proporción de bolas negras que en la saca B, pero ¡cuidado!, no que haya la misma proporción de bolas blancas y negras).

Para atrevernos a apostar juiciosamente, en el ejemplo antedicho (en el que en la saca A hay más bolas blancas que en la saca B, lo que también se puede describir como la afirmación de que en la saca B hay más bolas negras que en la saca A) necesitamos disponer de algún medio para, antes de apostar, calcular cuál es la probabilidad de sacar tres bolas blancas en A y dos bolas blancas en B, siendo que en A no hubiera más bolas blancas que en B.

Si esa probabilidad fuera pequeña, probablemente nos atreveríamos a apostar más fuerte (o más dinero) que si es algo de fácil ocurrencia por casualidad.

Una manera de estar más seguros, como sabemos, es aumentar el tamaño de la muestra. En efecto:

1) Si nuestra estimación procede de una muestra de cinco bolas, todos sabemos que, habiendo en la saca la mitad (50%) de cada color, no es demasiado difícil que salieran todas idénticas (o sea, el 100%). Concretamente, se puede calcular que la probabilidad es de dos veces entre 32 (es decir, una opción de que todas fueran blancas y otra de que fueran todas negras, entre las 32 alternativas posibles: $2/32 = 0,0625 = 6,25\%$).

2) Sin embargo, si la muestra es de 10 bolas, la probabilidad sería: $2/1024 = 0,2\%$.

3) Si, bajo las mismas condiciones (población de la saca: 50% blancas y 50% negras), la muestra fuera de 20 bolas, la probabilidad sería: $2/1048576 = 0,0002\%$.

O sea, al aumentar la muestra, se reduce exponencialmente la probabilidad (o sea, la esperanza matemática) de que las cosas raras ocurran. Y lo mismo sucede para cualquier combinación.

Por fortuna, las matemáticas han desarrollado procedimientos de calcular muy atinadamente (unas veces exactamente, y otras muy aproximadamente) la probabilidad de ocurrencia de diferentes eventos. Esos cálculos están al alcance de cualquier persona titulada en ciencias, porque se han desarrollado aplicaciones informáticas específicas para estadística, que permiten automatizar fácilmente las operaciones necesarias, sin necesidad de ser matemático o estadístico, igual que para escribir con un procesador de textos no hace falta ser informático.

Digamos que el estudio estadístico de una investigación también es más que manejar diestramente una aplicación SPSS, SAS, Stata, etc.

1) Si se conocen los fundamentos metodológicos, la comunicación con un experto estadístico puede ser breve, fluida y provechosa, y la mayoría de las veces se limitará a una supervisión en el muestreo y en el análisis final de los datos, porque la lectura de los outputs estadísticos y su interpretación científica, que son tareas y responsabilidad del investigador, estarán a su alcance.

2) Pero si no se dominan esos fundamentos, la situación es sumamente difícil, y suele tener los resultados catastróficos que se atisban en la mayoría de las publicaciones “científicas”, que no pocas veces son acientíficas.

IV. RELACIÓN Y CAUSALIDAD

1º Condiciones para la causalidad determinista.

Existen diversas definiciones de causa y de causalidad, según diferentes filosofías y circunstancias. En una concepción lógico-filosófica determinista, se dice que la variable X es causa: específica, necesaria y suficiente, si cualquier valor de X corresponde de manera única y específica un determinado valor de Y.

2º Cuando no hay una situación determinista.

Sin embargo, la causalidad única y determinista es excepcional, pues lo habitual, especialmente en el mundo de la biología, es entre las variables que exista:

1. Multicausalidad (también llamada con menos rigor causalidad multifactorial).
2. Multiplicidad de efectos (a veces diferentes, según la intensidad de la variable causal que consideremos).
3. Diferentes modalidades o sentidos de los efectos (todo o nada, efectos linealmente proporcionales, efectos no linealmente relacionados, etc.).
4. Interacciones o interferencias, no siempre lineales (agregables entre diferentes causas), y así es conocido que, por regla general, dos antibióticos bactericidas se potencian mutuamente en sus efectos, dos antibióticos bacteriostáticos se suman, y un antibiótico bacteriostático interfiere desfavorablemente con un bactericida cuando se combinan.
5. Limitaciones vinculadas a todas las medidas:
 - 1) Unas veces porque carecemos de unidades apropiadas (¿cómo medimos la cantidad fumada por un enfermo?, ¿por la cantidad de tabaco, de alquitrán, de nicotina, etc.?, ¿cómo comparamos puros, pipas o cigarrillos?, ¿cómo comparamos el grado de apuramiento de las colillas), ¿cómo comparamos los cambios cuantitativos en el hábito tabáquico a lo largo del tiempo que pueden influir en la enfermedad, habida cuenta de que no es lo mismo haber fumado mucho o fumar ahora poco, que al revés?
 - 2) Otras porque, aun habiendo unidad y procedimiento de medida, toda medida lleva inexorablemente ligado un error o falta de exactitud (por errores personales, limitaciones procedimentales o técnicas, errores instrumentales, etc.), ligados a las carencias técnicas, a la imprecisión de los aparatos de medida o a la imprecisión de las expresiones numéricas (propagada o exacerbada a veces con los cálculos), etc.
6. Complejidad del sistema de variables relacionadas, que a veces excede nuestra capacidad de comprensión.
7. Limitación de las posibilidades de cálculo.

Condiciones para multicausalidad (que suele denominarse sin rigor causalidad multifactorial).

En la multicausalidad, es decir, aquella en que la variable dependiente es consecuencia de varias causas, hay varios criterios para asociar la causalidad, a veces no fáciles de desenmascarar, y a veces imposibles. Entre los criterios necesarios para asociar la existencia de causalidad, destacan los tres siguientes:

1. Existencia de relación. Si la variable X tiene alguna influencia en la variable Y, quiere decir que los distintos valores de X se corresponden con diferentes valores de Y, por lo que X e Y han de ser variables relacionadas, es decir, han de estar ligadas por:

1) Una asociación, que se llama así a la relación que involucra a alguna variable cualitativa, en la que cada categoría, clase o modalidad (que es el nombre que se da a las diferentes opciones o valores de las variables cualitativas) está ligada con los diferentes valores de la otra variable; o

2) Una covariación, que es como se llama a la relación entre los valores numéricos de variables cuantitativas.

Ahora bien, la existencia de relación es una condición necesaria pero no suficiente de la existencia de causalidad (causalidad \Rightarrow relación), porque las variables X e Y pueden simplemente no estar relacionadas entre sí, pero parecerlo porque son consecuencia de otra variable, que es la causa influyente sobre los valores de X y de Y. Esta variable se llama colateral, y no siempre se consigue desenmascararla.

2. La temporalidad. Implica que la causa tiene que preceder al efecto. A veces, en respuestas muy rápidas de la variable Y (consecuencia) a la variable X (causa) pueden parecer simultáneas, porque no se puede discriminar con exactitud cuál de los cambios fue anterior. Pero nunca la variable consecuencia puede preceder a la variable causa.

3. Ausencia de espuriedad, es decir, de terceras variables o variables de confusión, que parasitan las aparentes relaciones.

En los estudios experimentales se cuenta con la ventaja de que los individuos con los que se experimenta pueden ser asignados, tras aleatorizarlos, al grupo de intervención o al grupo de control, lo que teóricamente equilibra las probabilidades de que las posibles variables de confusión puedan actuar en un sentido o en el contrario, tanto en unos grupos como en los otros de la experimentación que se va a comparar.