CALEIDOSCOPIO
11
ENERO-JUNIO 2002

Consideraciones sobre la interpretación de las medidas.

El caso de los coeficientes de Gini y de correlación

FELIPE MARTÍNEZ RIZO

Departamento de Educación/UAA

INTRODUCCIÓN

Para entender las frecuentes discusiones que enfrentan a partidarios de las técnicas cuantitativas y cualitativas en el campo de la investigación en ciencias del hombre, debe tenerse en cuenta, entre otras cosas, la dificultad que muchas personas tienen para entender correctamente los conceptos subyacentes a los procedimientos que implican números y su tratamiento aritmético y estadístico.

En este trabajo se reflexiona primeramente sobre la razón de que dicha dificultad sea tan frecuente, que se atribuye al carácter relativamente reciente no sólo de las técnicas sofisticadas, sino de los procedimientos básicos de medición, los que tienen que ver con la manera misma en que se observan ciertos aspectos de la realidad, determinadas variables que podemos llamar de primer orden o *primarias*, en particular las que se usan en ciencias del hombre, como el sexo de las personas, su estado civil, su capacidad intelectual, su estatura y su peso o su edad.

Luego se analizan los casos de dos medidas de un nivel de abstracción superior que no se refieren a ese tipo de variables *primarias* como las que se acaban de mencionar, sino a otras, que podríamos llamar de segundo orden, porque tienen que ver con la manera en que se distribuyen o relacionan entre sí las anteriores. Se trata del coeficiente de Gini, que mide la igualdad o desigualdad de la distribución de un bien entre una población, y de los coeficientes de correlación, que expresan la medida en que dos variables se presentan asociadas, de forma que cuando una aumenta o disminuye la otra también lo hace. Pese a tratarse de variables de segundo orden, su importancia puede apreciarse con facilidad si notamos el temprano momento en que ambas medidas se desarrollan: el coeficiente de Gini desde 1912; los de correlación aún antes, desde fines del siglo XIX.

Debe subrayarse que este artículo no pretende dar una aportación original a la teoría estadística; las propiedades matemáticas de las medidas mencionadas han sido analizadas detalladamente por los especialistas, pese a lo cual muchos investigadores de las ciencias sociales siguen desconociéndolas, con la consecuencia frecuente de usos inadecuados. Sin desconocer la importancia de una mejor formación matemática, que permita un manejo riguroso de los procedimientos estadísticos, en este trabajo se argumenta que es importante también la familiaridad con una medida y los resultados que arroja, para hacer interpretaciones adecuadas de ellos.

El texto es una reflexión de un investigador sin formación estadística sofisticada, pero con experiencia, que se dirige a otros investigadores con igual deficiencia en su formación matemática, pero con menos experiencia, para mostrar que aún sin la formación matemática deseable es posible evitar errores graves, de los que la sofisticación matemática, por sí sola, tampoco exime.

La observación de variables *primarias* y sus niveles de medición

Los investigadores en ciencias del hombre suelen estar familiarizados con las nociones relativas a los niveles de medición de las variables que utilizan, que la teoría distingue con los conocidos niveles: nominal, ordinal, de intervalo y de razón, si bien, solamente coinciden con la noción habitual de medir los dos últimos, que se expresan con la clase de números que los viejos libros de texto llamaban cardinales, los que se utilizaban para expresar la *cantidad* que, según esos mismos libros, se definía como *aquello que puede aumentar o disminuir, y que se puede contar, pesar o medir.*

En efecto, medir a nivel nominal no es otra cosa que lo que normalmente llamamos *clasificar*, como cuando agrupamos a las personas según su sexo o su estado civil; medir a nivel ordinal es lo mismo que *jerarquizar* u *ordenar*, como cuando un grupo de personas forma una línea en orden creciente o decreciente de estatura. Cuando tomamos la temperatura de un enfermo con un termómetro, en cambio, medimos a nivel de intervalo; y cuando estimamos la longitud de un objeto con una cinta métrica, medimos a nivel de razón. Es significativo que sea sólo en el último caso cuando se emplea normalmente el verbo *medir: medimos* el tamaño del objeto con una *cinta de medir.*

Los seres humanos nacemos equipados con una sofisticada capacidad para el lenguaje oral y con habilidades numéricas básicas, que permiten hacer conteos y operaciones aritméticas elementales con cantidades reducidas. No ocurre así con la habilidad de escritura y la de efectuar operaciones matemáticas complejas que deben aprenderse como productos culturales que son. Las civilizaciones que desarrollaron la escritura vieron surgir también la matemática, incluyendo la creación de sistemas de medición, si bien, inicialmente limitados a aspectos de la realidad que pueden medirse con relativa facilidad en los niveles de intervalo o de razón: el tamaño de los objetos, su volumen y su peso, o las distancias entre un lugar y otro: codos, pies, brazas y pulgadas; ánforas o pellejos; libras y quintales; estancias, estadios y leguas, son ejemplos de unidades que se definieron a partir de partes del cuerpo, recipientes y prácticas habituales, y luego se estandarizaron dando lugar a procesos estrictos de medición, formalmente idénticos a los actuales, si bien los procedimientos utilizados podían ser mucho menos precisos que estos últimos. Debe observarse que el nivel ordinal de medición no implica tales operaciones (definición de unidades estándar divisibles con las cuales comparar los objetos a medir), sino únicamente la comparación de los objetos entre sí, en tanto que el nivel nominal supone solamente las operaciones mentales de categorización y clasificación.

En forma bastante aproximada se medía también el tiempo. Otra realidad, que iría cobrando cada vez más importancia, al grado de que tal vez haya sido el factor que más influyó en el desarrollo de los sistemas de numeración, fue el dinero, en sus formas iniciales, como unidades para facilitar el trueque: cabezas de ganado (*pecus*, de donde viene *pecunia*, dinero), granos de cacao, etc. La medición de aspectos de la realidad menos *tangibles*, en cambio, es algo mucho más reciente: sólo a partir del siglo XVII, con el surgimiento de la ciencia moderna, se inventaron procedimientos para medir variables conocidas de tiempo atrás, como la temperatura, y se desarrollaron conceptos nuevos, que también fueron objeto de mediciones, como el de presión atmosférica.

La medición de intervalo y de razón no es, pues, una actividad *natural*, para la que el hombre esté programado evolutivamente, sino una *cultural*, que las civilizaciones desarrollan y perfeccionan a lo largo del tiempo, como muestra el que haya sido hasta principios del siglo XX cuando se desarrolló una teoría de la medición en general, y hasta la segunda mitad del mismo cuando se extendió ampliamente a las ciencias del hombre. Ese carácter no natural de la medición, y lo reciente de su sistematización, explican seguramente la dificultad que muchas personas experimentan para comprender todo lo relacionado con ella, y para manejarla con soltura y correctamente.

El manejo práctico de medidas conocidas

La dificultad de entender y manejar bien una medida se acentúa si se está poco familiarizado con ella. La comprensión de una medida, en efecto, puede alcanzarse de dos maneras. Una es por la vía que podemos denominar teórica y que consiste en comprender las características formales de una medida, sus propiedades matemáticas, los principios en que se basa su construcción; la otra es la que llamaremos práctica y se adquiere gracias a la utilización frecuente de la medida en casos diversos, lo que permite al usuario familiarizarse con ella aprendiendo su significado práctico que se concreta en cosas como los valores extremos que puede tomar en la vida real, lo relevante o irrelevante de cambios en cierta escala, etc.

La escasa facilidad espontánea para los números y la precaria formación matemática de la mayoría de la gente, hace que pocos alcancen un alto grado de familiaridad con la medición por vía teórica; muchas personas, en cambio, pueden adquirirla por vía práctica, por lo menos en ciertos ámbitos. Es conocida la soltura con que calculan precios y dan cambio quienes se dedican al comercio. Los dos ejemplos siguientes mostrarán que la mayoría de las personas maneja ciertas medidas con soltura suficiente para todos los efectos necesarios en la vida práctica.

- Primer ejemplo: si se toma la temperatura corporal a una persona y se le informa que tiene 41°, eso bastará para que sepa que está enfermo de cierta consideración, si no bastara para ello el malestar que indudablemente experimenta. Aunque no sepamos cómo se construyeron las escalas de temperatura, ni podamos convertir grados Celsius a Farenheit, muchos sabemos que deberíamos tener 36.5° o 37°, que incluso 38° es ya indicio de que algo no anda bien en nuestro organismo y que más de 40° es francamente para preocuparse.
- Segundo ejemplo: si nos dicen que la estatura promedio de un grupo de personas es de algo más de dos metros, muchos pensaremos que se trata de jugadores de un equipo profesional de baloncesto o, de cualquier modo, de un grupo sumamente atípico. Para ello no es necesario saber que el sistema métrico decimal es un fruto de la modernidad, y concretamente de la

Revolución Francesa, ni que la vieja definición del metro como la diezmillonésima parte del cuadrante del meridiano terrestre dejó el sitio hace dos siglos a la de la distancia entre dos marcas en una barra de platino iridiado que se guarda en una bóveda de seguridad en alguna parte del mundo, ni que en las últimas décadas ha sido substituida, a su vez, por una esotérica definición en términos de la longitud de onda de ciertas radiaciones. Para hacer la conjetura mencionada arriba basta tener una idea de que la estatura media de los varones adultos, en nuestro contexto, debe situarse alrededor de 1.70 m.

El caso del índice de Gini

Las medidas desarrolladas en las ciencias del hombre son relativamente recientes, pues las primeras datan sólo de fines del siglo XIX; la escasa difusión de los trabajos de estas disciplinas hace que en general se tenga poca familiaridad con ellos; además, pocas personas, incluso investigadores, dominan sus fundamentos teóricos; por todo ello la capacidad de comprensión y manejo adecuado de las medidas en las ciencias del hombre suele ser muy reducida, como muestra el ejemplo siguiente, que se refiere a una de las más antiguas que se sigue utilizando mucho, si bien en ámbitos muy especializados, como el de los estudios de distribución del ingreso: se trata del coeficiente o índice desarrollado en 1912 por el italiano Corrado Gini para medir el grado de desigualdad de la distribución de un bien entre los individuos de una población.

En la campaña electoral de 2000, un candidato afirmó que los grados de escolaridad estaban peor distribuidos que los ingresos monetarios. En apoyo de su afirmación citaba unos datos, según los cuales el 20% más pobre de los mexicanos recibe alrededor del 4% del total del ingreso nacional, en tanto que le corresponde solamente alrededor del 3% del total de grados de escolaridad acumulados por la población. Las cifras anteriores son aproximada-mente verdaderas, pero la afirmación general no lo es porque considera sólo a una quinta parte de la población, e ignora al 80% restante.

La distribución de los ingresos difiere de la distribución de la escolaridad en sus dos extremos: en el inferior, aunque las personas más pobres reciban algunos ingresos se registra a cierto número de ellos con ingresos nulos, pero sólo alrededor del 5%; en cuanto a la escolaridad, en cambio, todavía un 10% de los mayores de 15 años son registrados con educación cero: la población analfabeta. En el otro extremo, las distribuciones de educación e ingreso difieren también: mientras los más ricos reciben cantidades astronómicas, decenas o centenares de veces superiores a las que reciben los más pobres, la escolaridad tiene un límite superior relativamente cercano: pocas personas llegan a terminar un doctorado (lo que representa unos 20 a 22 años de estudios) y un número insignificante acumula algunos años más de escolaridad formal, o sea sólo tres o cuatro veces más que quienes únicamente hayan terminado la primaria.

En una sociedad que logre universalizar la educación elemental, pues, la escolaridad debería estar significativamente *mejor* distribuida que los ingresos monetarios, y no peor que ellos, sobre todo en el caso de sociedades con altos niveles de desigualdad económica. Por tales características de las distribuciones de distintos bienes, para comparar unas con otras se necesitan medidas de desigualdad que consideren todo el rango de la población, y no sólo parte de ella, como se hizo en la comparación de la campaña electoral. El índice de Gini es una medida así y, pese a su antigüedad, es una de las más utilizadas.[1]

---

[1] *Cfr.* Fernando Cortés y Rosa Ma. Rubalcava, *Técnicas estadísticas para el estudio de la desigualdad social*. México, El Colegio de México, 1982; y Philip B. Coulter, *Measuring Inequality. A Methodological Handbook*. Boulder, Co., Westview Press, 1989.

Un campo en que se maneja el índice de Gini es, justamente, el análisis de la desigualdad educativa. Por ello llama la atención que, en la década de 1970, en América Latina se aceptara ampliamente una idea similar a la que se manejó en la campaña presidencial mexicana treinta años después, la de que la educación estaba más mal distribuida que el ingreso. Retomando un análisis publicado previamente,[2] observamos que, al iniciar la década mencionada, un texto sobre la educación en Colombia[3] llegaba justamente a la conclusión en cuestión, comparando el índice de Gini de la distribución del ingreso en aquel país (0.54) con los valores del mismo índice aplicado a la distribución de la escolaridad, que arrojaba valores altísimos, llegando a 0.95, según los cálculos del investigador.

Si sólo se considera que el índice está construido de forma que los valores extremos que puede adoptar son 0 (desigualdad nula, igualdad absoluta) y 1 (desigualdad absoluta), y si no se ha usado en varios contextos, adquiriendo un mínimo de familiaridad con la medida, es fácil aceptar una cifra tan alta sin sospechar que pueda haber algún error. Pero revisando otros estudios sobre desigualdad educativa puede verse que, si bien los extremos del índice de Gini son teóricamente 0 y 1, en la práctica, sistemas educativos bastante igualitarios no tienen índices menores a 0.15, en tanto que situaciones de desigualdad grave se reflejan en valores del orden de 0.90.

---

2 Felipe Martínez Rizo, "La desigualdad educativa en México, 1970-1990", *Revista Latinoamericana de Estudios Educativos*. Vol. XXII (1992) N° 2, pp. 59-120.

3 M. Urrutia Montoya, "Distribución de la educación y distribución del ingreso en Colombia", *Revista del Centro de Estudios Educativos*. Vol. II (1972) N° 1, pp. 32-59.

El índice de Gini de la distribución de la escolaridad en Estados Unidos en 1960, por ejemplo, era de 0.17; entre la población económicamente activa del Reino Unido en 1951 era de 0.18; y en la población masculina de 25 años o más del Japón, en 1960, de 0.24. En el otro extremo, algunos de los valores más altos llegan a 0.92 (Túnez, 1966); 0.87 (Argelia, 1971); 0.80 (Kenia 1969 y Guatemala 1957) y 0.78 (India 1971 y Brasil 1950).[4]

Una cifra de 0.95, por consiguiente, es sumamente alta, y no parece verosímil que corresponda a una situación como la de Colombia, y menos a la población de 20 a 24 años de la región de Bogotá, como aseguraba Urrutia Montoya. Al revisar los cálculos pudo detectarse la presencia de un error que los afectaba sistemáticamente, y al recalcular los índices se encontraron valores del orden de 0.5, esperables e inferiores a los índices de distribución del ingreso. Sin un análisis matemático profundo, el conocimiento de los valores habituales del índice en varios contextos permitió detectar un error sistemático que había pasado desapercibido a investigadores que tal vez contaban con mejor formación, pero no estaban familiarizados con el tema.[5]

El caso de los coeficientes de correlación

Algo similar ocurre con otras medidas antiguas, los coeficientes de correlación, como el llamado de producto-momento, desarrollado a fines del XIX por Karl Pearson. La importancia de la noción misma de correlación que, con sus múltiples variantes, está en la base de la estadística multivariada, hace que su compren-

4 M. Urrutia Montoya, *op. cit.*, p. 43; George Psacharopoulos, y M. Woodhall, *Educación para el desarrollo. Un análisis de opciones de inversión*. Madrid, Tecnos, 1987, pp. 234-235.

5 *Cfr.* Martínez Rizo, *op. cit.*

sión parezca fundamental para los investigadores sociales, pero la realidad parece ser, por el contrario, que una proporción preocupantemente alta de ellos no tiene ni un dominio formal de las propiedades matemáticas de la medida, ni una familiaridad práctica con ella, de manera que pocos parecen estar en condiciones de entenderla y manejarla correctamente.

Seguramente lo anterior explica que algunos textos propongan interpretaciones convencionales de los coeficientes de correlación, que pueden ser útiles como una primera aproximación, pero deben utilizarse con múltiples salvedades. Guilford, por ejemplo, propone las siguientes interpretaciones:[6]

CUADRO 1

INTERPRETACIÓN DEL COEFICIENTE R DE PEARSONS

Valores de r	*Interpretación descriptiva*
Menor que 0.20	Correlación leve, casi insignificante
De 0.20 a 0.40	Baja correlación, definida pero baja
De 0.40 a 0.70	Correlación moderada, sustancial
De 0.70 a 0.90	Correlación marcada, alta
De 0.90 a 1.00	Correlación altísima, muy significativa

En forma similar, Nelson propone interpretaciones convencionales para el coeficiente de correlación φ (phi), que fue desarrollado también por Pearson para el caso de variables nominales.[7]

6 En Jorge Padua, *Técnicas de investigación aplicadas a las ciencias sociales*, México, FCE, 1979, p. 287.

7 En Mohammad Naghi Namakforoosh, *Metodología de la investigación*, México, Limusa, 1988, p. 308.

CUADRO 2

INTERPRETACIÓN CONVENCIONAL DEL COEFICIENTE φ

Valor absoluto de la correlación phi	*Interpretación de la relación*
Arriba de 0.80	Muy fuerte
0.60 a 0.80	Más o menos
0.40 a 0.60	Débil
0.20 a 0.40	Muy débil
0.00 a 0.20	No existe

Los coeficientes de correlación suelen construirse de manera que sus valores extremos sean 0 y 1, equivaliendo el 0 a ausencia total de relación y el 1 a relación perfecta, directa (+1) o inversa (–1). Por ello una correlación de 0.5 se interpreta espontáneamente como *media*; una de 0.75 como *significativa* en el sentido no técnico de *bastante fuerte, pero no demasiado*; y una de 0.25 como bastante débil, pero no insignificante. Estas interpretaciones suelen hacerse sin las precauciones que las características de los coeficientes aconsejan, como verificar la linearidad de la relación, someter a prueba la bondad del ajuste, tener en cuenta el tamaño de la muestra, la presencia de normalidad, etc. Si no se tiene conocimiento de diversos casos en que se utilice un coeficiente, será fácil cometer serios errores, tanto más graves cuanto que la falta de cultura estadística y de familiaridad con el uso de técnicas cuantitativas, tan frecuentes en los medios de investigación educativa, hacen que los errores pasen desapercibidos y que interpretaciones carentes de sustento se acepten como buenas.

Suele afirmarse que los coeficientes r y φ desarrollados por Pearson para variables métricas y nominales, respectivamente, son equivalentes. Lo son, en efecto, si se trata de variables dicotómicas que adopten los valores 0 y 1. Pero no cuando una variable métrica, que adopta realmente diversos valores, se hace artificialmente dicotómica. Obviamente, al hacer lo anterior se pierde información, y el coeficiente j que se calcula sobre los valores resultantes puede arrojar valores muy *inferiores* a los que adopta el coeficiente r calculado sobre los valores originales.

La pérdida de información de una medida apropiada para variables nominales, como el coeficiente φ cuando se usa para variables medidas en un nivel superior, artificialmente reducidas a un nivel nominal, puede hacer que φ tome valores inferiores a los que adoptan las medidas apropiadas para variables de intervalo o de razón, como el coeficiente r, como se mostrará más adelante. Ésta es una de las consideraciones prácticas que habrá que tomar en cuenta para interpretar los valores que adopta en una investigación un coeficiente determinado.

Otra consideración es la que se refiere a que en las ciencias del hombre es difícil encontrar correlaciones tan elevadas como las que son normales, en algunos casos, en las ciencias de la naturaleza, de suerte que un valor de 0.5, que en las segundas puede considerarse medio, en las primeras, en ciertas circunstancias, puede ser bastante elevado. En los estudios del valor predictivo de pruebas como las que se utilizan para selección de alumnos en instituciones americanas de educación superior, por ejemplo, los coeficientes de correlación entre los puntajes obtenidos por los aspirantes en las pruebas y el promedio de las calificaciones obtenidas por los aceptados en su primer año de estudios difícilmente alcanzan el valor de 0.6, incluso si se aplica la corrección llamada por restricción de rango, que hace subir los coeficientes originales, normalmente ubicados en el orden de 0.5, poco más o menos. Si se sabe lo anterior, un coeficiente de 0.45, obtenido en un estudio de valor predictivo de una prueba mexicana, no será considerado inadecuado, como sucede con facilidad si se ignora ese tipo de datos, y se aplica de manera simplista una tabla de interpretaciones convencionales como las dos anteriormente citadas.

LA INTERPRETACIÓN DE LOS COEFICIENTES PARA VARIABLES REGISTRADAS A DIFERENTE NIVEL DE MEDICIÓN

El ejemplo siguiente muestra la reducción del valor de los coeficientes, si se pasa de uno apropiado para un nivel de medición más alto a uno más bajo, construido artificialmente a partir de los datos del nivel superior de medición.

Si se calcula el coeficiente r de Pearson para un conjunto de datos medidos en el nivel de intervalo, como los de estatura y peso de diez sujetos que contiene el cuadro 3, y luego se calculan los coeficientes ρ de Sperman (apropiado para medidas ordinales) y el ya mencionado φ (adecuado para datos medidos en el nivel nominal) para el mismo grupo de datos, reduciendo previamente su nivel de medición, se observará cómo el valor del coeficiente disminuye. Como las dos variables están medidas originalmente en un nivel de razón, es apropiado el coeficiente r de Pearsons para medir la correlación entre ambas; al calcularlo se obtiene el valor de 0.690.

CUADRO 3

ESTATURAS Y PESOS DE 10 SUJETOS

Sujetos	*Estatura en cms.*	*Peso en kgs.*
1	155	48
2	156	45
3	162	52
4	167	56
5	160	57
6	162	56
7	158	47
8	159	59
9	168	59
10	163	54

A partir de una medida de nivel superior es fácil obtener una de nivel inferior; en este ejemplo es sencillo obtener una medida ordinal de estaturas y pesos de los sujetos del cuadro, identificando el lugar o rango que ocupa cada uno en el conjunto en relación con las dos variables, como se hace en el cuadro siguiente, en el que al sujeto 1, el más bajo en estatura, se asigna el lugar 10° en ese aspecto, en tanto que el 10° lugar en peso corresponde al sujeto 2, que pesa sólo 45 kg. Cuando hay empates se asigna un valor intermedio a los sujetos. Calculando el coeficiente ρ de correlación de rangos de Sperman, que es apropiado para medidas ordinales, se obtiene un valor de 0.594.

CUADRO 4

LUGAR QUE OCUPAN LOS 10 SUJETOS, SEGÚN ESTATURA Y PESO

Sujetos	*Lugar en estatura*	*Lugar en peso*
1	10°	8°
2	9°	10°
3	4.5°	7°
4	2°	4.5°
5	6°	3°
6	4.5°	4.5°
7	8°	9°
8	7°	1.5°
9	1°	1.5°
10	3°	6°

A partir de los datos de cualquiera de los dos cuadros anteriores podemos obtener ahora uno más, en el que los sujetos se clasifiquen simplemente en forma dicotómica, nominal, según que se sitúen en la mitad de individuos más altos o más bajos, más pesados o más livianos.

CUADRO 5

CLASIFICACIÓN DICOTÓMICA DE LOS SUJETOS POR ESTATURA Y PESO

Sujetos	*Clasificación por estatura*	*Clasificación por peso*
1	Bajo	Liviano
2	Bajo	Liviano
3	Alto	Liviano
4	Alto	Pesado
5	Bajo	Pesado
6	Alto	Pesado
7	Bajo	Liviano
8	Bajo	Pesado
9	Alto	Pesado
10	Alto	Liviano

A partir del cuadro 5 se puede construir fácilmente el 6, cuyas cuatro celdillas centrales reflejan el cruce de las dos variables dicotomizadas, y a partir del cual se calcula el coeficiente φ que es uno de los derivados de la χ^2 según la fórmula siguiente: $\varphi = \sqrt{\varphi^2} = \sqrt{\chi^2 / N}$

CUADRO 6

CLASIFICACIÓN DICOTÓMICA POR ESTATURA Y PESO

		Sujetos según estatura		*Totales de*
		altos	*bajos*	*renglones*
SUJETOS SEGÚN PESO	PESADOS	3	2	5
	LIVIANOS	2	3	5
TOTALES DE COLUMNAS		5	5	10

Los valores que corresponden al cuadro anterior son: $N = 10$; $\chi^2 = 0.4$; $\varphi^2 = 0.04$; $\varphi = 0.2$.

Puede obtenerse también el coeficiente C de Pearson, teóricamente equivalente a r, y que se define por la fórmula $X = \sqrt{\chi^2 / N + \chi^2}$. En este caso se obtiene el valor de C = 0.196.

Puede apreciarse cómo se reducen los valores de los coeficientes r, ρ, φ y C, aunque corresponden al mismo conjunto de datos: la estatura y el peso de 10 sujetos; inicialmente se trata de variables métricas, medidas a nivel de razón, de las que se obtiene r; después de transformarlas en variables de niveles ordinal y nominal, los coeficientes correspondientes adoptan valores inferiores, como se resume en el siguiente cuadro.

CUADRO 7

COMPARACIÓN DE LOS TRES COEFICIENTES DE CORRELACIÓN

Coeficiente de Correlación	*Valor obtenido en el ejemplo*
r de Pearson	0.690
ρ de Sperman	0.594
φ de Pearson	0.200
C de Pearson	0.196

La reducción observada de los valores de los coeficientes se explica posiblemente por la pérdida de información que ocurre al bajar de r a ρ y luego a φ y muestra que las recomendaciones del cuadro 2, en el que se propone interpretar como inexistente una correlación de 0.2 con el coeficiente φ deben interpretarse con cuidado, distinguiendo el caso de variables realmente nominales, como el género, de las variables que podrían medirse a un nivel superior, ordinal o incluso métrico, y que por limitaciones de los procedimientos de recolección de datos se manejan como nominales.

Obviamente no es lógico bajar de nivel de medición un conjunto de datos que ya se ha obtenido con un nivel más exigente. Pero lo que ocurre a veces es que, por la dificultad que representa diseñar y utilizar instrumentos de medición rigurosos para captar variables usuales en las ciencias del hombre, como actitudes u opiniones, un investigador utilice coeficientes apropiados a un nivel nominal e interprete los resultados sin fijarse en que los valores resultantes serán menores a los que podría obtener si manejara instrumentos que le permitieran tener variables métricas, y luego empleara un coeficiente de correlación para ese nivel. El error opuesto es el que consiste en utilizar un coeficiente para variables métricas cuando sólo se cuenta con datos ordinales o nominales.

Para explorar la interpretación de la reducción del valor de los diversos coeficientes de correlación, como un efecto de la pérdida de información que ocurre al bajar una variable de nivel de medición, haremos una variante del ejercicio anterior, pasando nuevamente las variables de estatura y peso al nivel nominal, pero ahora en forma tricotómica: utilizaremos los valores de alto, medio y bajo en las dos variables, clasificando a las tres personas más altas o pesadas en el valor alto, las cuatro siguientes en el valor medio, y las tres últimas en el valor bajo. Antes de hacer el ejercicio podemos hipotizar que el valor resultante deberá ser mayor al obtenido en el caso de dicotomización, con el coeficiente φ (0.2), puesto que la pérdida de información será menor.

CUADRO 8

CLASIFICACIÓN TRICOTÓMICA DE LOS SUJETOS POR ESTATURA Y PESO

Sujetos	*Clasificación por estatura*	*Clasificación por peso*
1	Bajo	Liviano
2	Bajo	Liviano
3	Mediano	Mediano
4	Alto	Mediano
5	Mediano	Pesado
6	Mediano	Mediano
7	Bajo	Liviano
8	Mediano	Pesado
9	Alto	Pesado
10	Alto	Mediano

A partir del cuadro 9, derivado del anterior, se puede calcular la correlación entre peso y estatura, utilizando ahora el coeficiente T de Tschuprov, que debe usarse en lugar de φ cuando no se trata de variables dicotómicas sino de tres o más valores, en tablas simétricas, o sea que tengan el mismo número de columnas y renglones, como es el caso, puesto que las dos variables se manejan como tricotómicas. Podría utilizarse también el coeficiente V de Kramer, que es adecuado para cualquier tipo de tabla, y en el caso de simétricas equivale a T, y en tablas de 2 x 2 a φ . La fórmula del coeficiente es $T = \sqrt{T^2} = \sqrt{\chi^2} / N \sqrt{(r-1)(c-1)}$

CUADRO 9

CLASIFICACIÓN TRICOTÓMICA POR ESTATURA Y PESO

		Sujetos según estatura			*Total*
		altos	*medios*	*bajos*	
SUJETOS SEGÚN PESO	PESADOS	1	2	0	3
	MEDIOS	2	2	0	4
	LIVIANOS	0	0	3	3
	TOTAL	3	4	3	10

Nota 1: Los valores que corresponden al cuadro anterior son: $N = 10$; $\chi^2 = 10.277$; $T^2 = 0.514$; $T = 0.717$.

Nota 2: El coeficiente C de Pearson ($C = \sqrt{c^2} / N + \chi^2$) en este caso es $C = 0.712$.

El cambio de un manejo dicotómico a uno tricotómico produce un incremento muy fuerte del coeficiente de correlación que alcanza un valor superior a los obtenidos previamente, no sólo con φ sino también con ρ e incluso con r, en una forma que coincide con las recomendaciones convencionales de interpretación de los coeficientes de correlación. (*cfr*. cuadros 1 y 2)

No es insignificante, por tanto, manejar una variable como dicotómica o como tricotómica, y la interpretación de los coeficientes de correlación debe tener en cuenta las posibles consecuencias de hacerlo de una u otra manera.

Para interpretar correctamente las cifras obtenidas en un estudio concreto debe considerarse, pues, si el nivel de medición de las variables corresponde efectivamente al que implica el coeficiente utilizado, además de verificar si se cumplen los demás supuestos que requiera cada medida, según sus propias características.

La sensibilidad del coeficiente r y su interpretación

El coeficiente r, apropiado para variables métricas, tiene la ventaja de que maneja un máximo de información, pero esto trae consigo la consecuencia de que r es muy sensible a factores como la normalidad de la distribución, o la presencia de valores extremos, de manera similar a lo que ocurre con las medidas de tendencia central y dispersión a partir de las que se construye r, la media y la desviación estándar.

En el caso de estudios de validez predictiva de pruebas de ingreso a instituciones de educación, como los mencionados anteriormente, por ejemplo, se obtendrán valores más elevados si las calificaciones asignadas a los alumnos admitidos se distribuyen en forma que se aproxima a una curva normal, por ejemplo con una mayoría de ochos, menos sietes y nueves y pocos dieces y seises. En cambio, en casos en que las calificaciones posteriores de los alumnos se apartan de la normalidad, por ejemplo cuando se encuentra un número desproporcionado de calificaciones muy elevadas (o muy bajas) como ocurre en algunas instituciones y carreras, la correlación es, obviamente mucho más baja: la correlación entre unos valores distribuidos normalmente y una constante es estrictamente igual a 0.

Una manera fácil de comparar situaciones diversas, aun sin tener una amplia experiencia real de investigación, consiste en construir artificialmente situaciones características imaginarias, y observar cómo se comportan en cada caso los coeficientes.

Esto es lo que se hace en los cuadros siguientes, en los que se presentan situaciones imaginarias de estudios de validez predictiva, con un grupo de sujetos que obtendría ciertos puntajes en la prueba de ingreso (que se supone normalizada, con una media de 1000 y una desviación estándar de 100) y el promedio obtenido por los mismos sujetos posteriormente, durante el primer año de sus estudios. Del análisis de los cuadros 10, 11 y 12 podemos derivar lecciones interesantes para una adecuada interpretación de los valores del coeficiente de correlación r.

CUADRO 10

Correlaciones entre puntajes en una prueba de admisión y promedio de calificaciones posteriores

		Promedio Posterior			
Sujetos	*Puntaje en prueba*	*Caso 1*	*Caso 2*	*Caso 3*	*Caso 4*
1. Abelardo	1300	10.00	06.00	08.00	10.00
2. Bernardo	1250	09.66	06.33	08.00	09.90
3. Carlos	1200	09.33	06.66	08.00	09.80
4. Daniel	1150	09.00	07.00	08.00	09.20
5. Enrique	1100	08.66	07.33	08.00	09.10
6. Francisco	1050	08.33	07.66	08.00	08.50
7. Gabriel	1000	08.00	08.00	08.00	08.20
8. Heliodoro	950	07.66	08.33	08.00	07.80
9. Isaías	900	07.33	08.66	08.00	07.60
10. Javier	850	07.00	09.00	08.00	07.40
11. Karl	800	06.66	09.33	08.00	07.10
12. Leonardo	750	06.33	09.66	08.00	06.70
13. Miguel	700	06.00	10.00	08.00	06.00
Media	1000	08.00	08.00	08.00	08.25
Correlaciones entre prueba y promedio		+ 1.00	- 1.00	0.00	+0.9925

Nota: Las cuatro columnas que se agrupan bajo el encabezado Promedio Posterior presentan otros tantos casos diferentes:

- En el primero, la correlación entre los puntajes obtenidos por los imaginarios sujetos en la prueba y su promedio posterior es positiva y perfecta: el puntaje en la prueba permite predecir perfectamente el rendimiento posterior: quien obtiene un puntaje más alto en la prueba obtendrá también, inexorablemente, mejores calificaciones.
- En el segundo caso, la correlación es también perfecta, pero negativa: quienes obtienen los puntajes más altos en la prueba obtienen, también invariablemente, las calificaciones posteriores más bajas.
- El tercer caso es uno de correlación nula, entre puntajes en la prueba que se distribuyen normalmente alrededor de la media de 1000, y unas calificaciones posteriores constantes, cuya media es de ocho, como en los dos casos anteriores, pero sin variación alguna.
- El cuarto caso es uno en el que los sujetos obtendrían en sus estudios posteriores notas un poco diferentes a las del caso 1, pero manteniendo la relación muy estrecha con los puntajes en la prueba; el coeficiente es prácticamente igual a 1.

El cuadro 11 presenta cuatro casos más, en los cuales hay un patrón semejante a uno de los del cuadro anterior, pero con casos discrepantes, más o menos graves, pero siempre excepcionales, que permiten apreciar cómo se modifica el coeficiente de correlación correspondiente.

CUADRO 11

CORRELACIONES ENTRE PUNTAJES EN PRUEBA DE ADMISIÓN Y PROMEDIO DE CALIFICACIONES POSTERIORES. CASOS IMAGINARIOS DE SUJETOS DISCREPANTES

		Promedio Posterior			
Sujetos	*Puntaje en prueba*	*Caso 1*	*Caso 2*	*Caso 3*	*Caso 4*
1. Abelardo	1300	10.00	10.00	08.00	06.00
2. Bernardo	1250	09.66	09.66	09.66	09.66
3. Carlos	1200	09.33	09.33	09.33	09.33
4. Daniel	1150	09.00	09.00	09.00	09.00
5. Enrique	1100	08.66	08.66	08.66	08.66
6. Francisco	1050	08.33	08.33	08.33	08.33
7. Gabriel	1000	08.00	08.00	08.00	08.00
8. Heliodoro	950	07.66	07.66	07.66	07.66
9. Isaías	900	07.33	07.33	07.33	07.33
10. Javier	850	07.00	07.00	07.00	07.00
11. Karl	800	06.66	06.66	06.66	06.66
12. Leonardo	750	06.33	06.33	06.33	06.33
13. Miguel	700	08.00	10.00	08.00	10.00
MEDIA	1000	08.15	08.30	08.00	08.00
Correlaciones entre prueba y promedio		+0.904	+0.623	+0.777	+0.209

- El primer caso del cuadro 11 es casi idéntico al primero del cuadro 10, con excepción del sujeto 13, que en el cuadro 10 presentaba un promedio de 6.00, el más bajo del grupo, coincidiendo con su puntaje de 700 en la prueba, igualmente el peor de los 13 sujetos; la correlación entre puntaje y promedio posterior era perfecta: 1.00. En cambio, aunque 12 de los 13 sujetos del primer caso del cuadro 11 mantienen la misma correlación y sólo el último se aparta de la tendencia, el coeficiente de correlación baja 10 centésimas y queda en 0.90.

- El segundo caso presenta una situación en la que 12 de los sujetos tienen puntajes y promedios idénticos a los del primer caso del cuadro 10, y nuevamente sólo uno, el último, se aparta de la situación anterior, ahora en forma extrema, pues el sujeto 13, con el peor puntaje en la prueba, coincide con el 1 en tener el promedio más alto del grupo; el resultado es que el coeficiente r, con un solo caso discrepante extremo, baja casi 40 centésimas, ubicándose en 0.623.
- El tercer caso presenta 11 de los 13 casos con valores en prueba y promedio idénticos a los del caso 1 del cuadro 10, con dos sujetos moderadamente discrepantes, el 1 y el 13; en este caso el coeficiente de correlación baja a 0.777.
- El último caso del cuadro 11 presenta igualmente una situación en la que los sujetos 1 y 13 tienen situaciones discrepantes, pero en este caso ambas extremas, con el mejor sujeto en la prueba como el peor en el promedio posterior, y viceversa. En este caso, el coeficiente de correlación baja hasta 0.209, cifra que, según las interpretaciones convencionales del cuadro 1 debería considerarse casi insignificante; al ver los valores de los que surge, sin embargo, no parece que tal interpretación pueda sostenerse sin matices. La conocida sensibilidad a los valores extremos de la media y la desviación estándar, es también, pues, una característica del coeficiente r, pariente cercano de esas medidas de tendencia central y dispersión, como se ha apuntado antes.

Un razonamiento similar al que se presenta aquí es el que hacen Tirado y colaboradores,[8] al explicar el sentido del coeficiente de correlación, precisamente en el marco de un trabajo

---

8 Felipe, E. Tirado,. Backhoff, N. Larrazolo y M. Rosas, "Validez predictiva del Examen de Habilidades y Conocimientos Básicos", en *Revista Mexicana de Investigación. Educativa,* Vol. II (1997) N° 3, pp. 67-84.

sobre la validez predictiva de la prueba de admisión desarrollada por ellos. Con ejemplos imaginarios, similares a los de los cuadros anteriores, esos autores muestran cómo:

- La correlación entre las puntuaciones en el examen de admisión y el promedio de las calificaciones de un alumno en el primer año de estudios posteriores es apenas superior a 0.5 si el 10% inferior de los sujetos presenta una situación anómala extrema, con un puntaje igual al del 10% más alto (ejemplo 1).
- Algo similar ocurre si un 15% en las dos puntas de la distribución presenta una situación anómala, aunque no extrema (ejemplo 2).
- Y si el 10% en las dos puntas presenta situaciones anómalas extremas (ejemplo 3) el coeficiente se reduce casi a 0, pese a que, como subrayan los autores, el 80% restante de la población presente una correlación perfecta.

CUADRO 12

EJEMPLOS HIPOTÉTICOS DE CORRELACIÓN ENTRE DOS MEDICIONES

Ejemplo 1			*Ejemplo 2*			*Ejemplo 3*		
SUJETOS %	ACIERTOS EXAMEN ADMISIÓN	PROMEDIO ESCOLAR AL 1ER AÑO	SUJETOS %	ACIERTOS EXAMEN ADMISIÓN	PROMEDIO ESCOLAR AL 1ER AÑO	SUJETOS %	ACIERTOS EXAMEN ADMISIÓN	PROMEDIO ESCOLAR AL 1ER AÑO
10	10	100	10	10	100	15	10	50
10	20	20	10	20	20	5	20	20
10	30	30	10	30	30	10	30	30
10	40	40	10	40	40	10	40	40
10	50	50	10	50	50	10	50	50
10	60	60	10	60	60	10	60	60
10	70	70	10	70	70	10	70	70
10	80	80	10	80	80	10	80	80
10	90	90	10	90	90	5	90	90
10	100	100	10	100	10	15	100	50
Correlación = 0.536			Correlación = 0.018			Correlación = 0.582		

Fuente: Tirado *et al.*, *op. cit.*, p. 73, Tabla 1.

La sensibilidad de r a valores extremos hace recomendable el empleo de variantes que los eliminen como forma de hacer más robusta la medida, de manera análoga a las medidas de tendencia central y dispersión que propone el enfoque de Análisis Exploratorio de Datos (*Exploratory Data Analysis, EDA): midmean, trimmed mean, winzorized mean, best easy systematic estimate* y similares.

UNA APROXIMACIÓN DIFERENTE A LA INTERPRETACIÓN DE LA CORRELACIÓN

Una manera diferente de abordar la cuestión es la que presentan Rosenthal y Rubin.[9] A partir de coeficientes de correlación encontrados en estudios sobre la efectividad de ciertos tratamientos médicos sobre algunas enfermedades, estos autores presentan las probabilidades de sobrevivir sin tomar la medicina o tomándola. Los resultados se resumen en el cuadro siguiente.[10]

---

9 Robert Rosenthal y Donald B. Rubin, "A simple, general purpose display of magnitude of experimenter effect", en *Journal of Educational Psychology*, N° 74, 1982, pp. 166-169.

10 *Cfr.* Frank J. Sulloway, *Born to Rebel. Birth Order, Family Dynamics, and Creative Lives*, New York, Pantheon Books, 1996. Appendix 1. A brief introduction to statistics (or correlations made easy), p. 372, Tabla N° 8.

CUADRO 13

SIGNIFICADO TÉCNICO Y PRÁCTICO DE LOS COEFICIENTES DE CORRELACIÓN

Significado Técnico			*Significado Práctico*		
COEFICIENTE CORRELACIÓN (r)	DESCRIPCIÓN CUALITATIVA DE LA CORRELACIÓN	VARIANZA EXPLICADA (r^2)	PROBABILIDADES DE SOBREVIVIR SIN MEDICINA	PROBABILIDADES DE SOBREVIVIR CON MEDICINA	MEJORA PROBABILIDADES SIN MEDICINA
0.01	Muy chica	0.01%	49.5%	50.5%	2%
0.05	Chica	0.25%	47.5%	52.5%	11%
0.10	Chica	1%	45%	55%	22%
0.20	Moderada	4%	40%	60%	50%
0.30	Moderada	9%	35%	65%	86%
0.40	Grande	16%	30%	70%	133%
0.50	Grande	25%	25%	75%	200%
0.80	Muy grande	64%	10%	90%	800%

Comentando los datos del cuadro anterior Sulloway dice:

La Tabla permite apreciar que incluso correlaciones pequeñas tienen efectos significativos. Una correlación de 0.10, por ejemplo, equivale a una mejora de las probabilidades de sobrevivir una enfermedad potencialmente fatal que pasa, de 45%, en caso de no tomar el medicamento, al 55% en caso de tomarlo. Esta mejora representa un incremento de las probabilidades de supervivencia del 22% con respecto a la probabilidad inicial, sin tomar medicina (55/45=1.22). Si su vida estuviera amenazada por una enfermedad que fuera potencialmente fatal, Ud. probablemente agradecería la posibilidad de tener acceso a una medicina de la que se supiera que tenía un efecto benéfico, inclusive con una correlación tan moderada. Una correlación de 0.30 equivale a casi duplicar las probabilidades de sobrevivir, que pasan de 35% a 65%. Una correlación de 0.50 produce una diferencia aún más notable de las probabilidades de sobrevivir, que pasan de 25% a 75%, una ganancia de 200% o, lo que es lo mismo, el triple de las probabilidades de supervivencia sin tomar la medicina.[11]

---

11 *Ibid.*, pp. 371-372.

Conclusión

Para finalizar podemos señalar que hay que tener cuidado al interpretar una medida, trátese del índice de Gini, un coeficiente de correlación, o cualquiera otra. Estrictamente hablando, esta recomendación es una obviedad; la experiencia muestra, sin embargo, que no está de más, sobre todo en un medio con escasa tradición en el campo de las investigaciones de corte cuantitativo.

Además de analizar las propiedades matemáticas de una medida para saber cuándo es adecuado utilizarla y verificar si se cumplen las condiciones pertinentes (la importancia de lo cual conviene enfatizar una y otra vez), es necesario adquirir un mínimo de familiaridad práctica con la medida, analizando trabajos que la utilicen, en contextos diversos, o mediante la construcción imaginaria de casos típicos que permitan ver, más allá de los valores que en teoría pueda tomar, el rango en el que puede moverse realmente, y nos permitan tener una idea de las consecuencias prácticas de ciertos valores. Para continuar con las referencias médicas, la formación estadística es muy valiosa, pero además hace falta la práctica, para desarrollar el *ojo clínico* que se necesita para la investigación. ❁