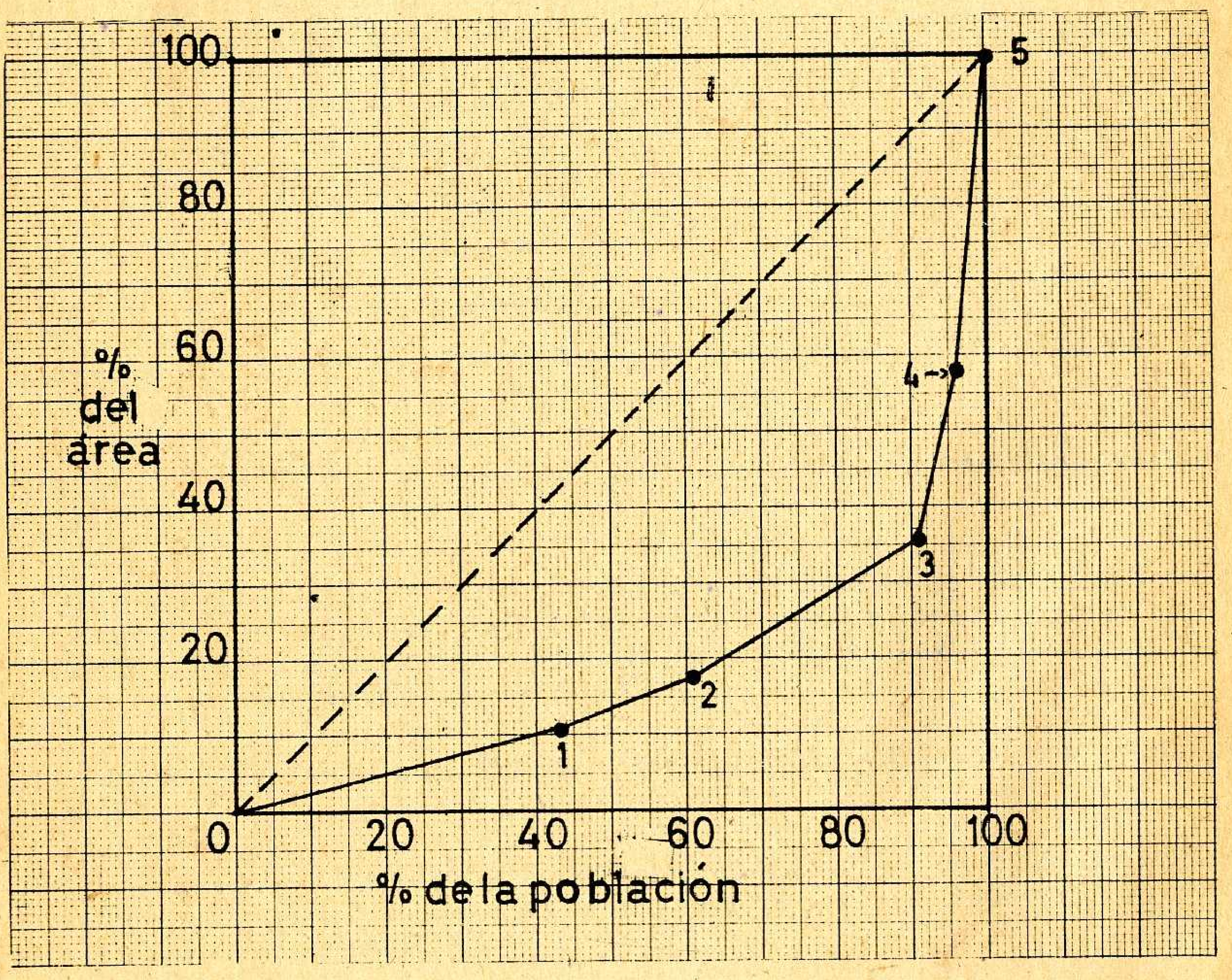
MATEMATICA EN LAS CIENCIAS SOCIALES

Por Lic. Luis Midence

En nuestro primer artículo(*), habíamos hablado de la curva de Lorenz, un instrumento muy útil para el investigador y para el educador. Al final del artículo, dejamos unas cifras para que los lectores pudieran practicar sus nuevos conocimientos. Las cifras correspondían a Brasil, y he aquí la gráfica de la curva de Lorenz y su coeficiente gini.

CURVA DE LORENZ PARA LA TABLA 2.0



(*) Magister, Año 1, No. 1, abril 1978.

Como se nota enseguida, la población de Brasil está muy concentrada en un área (el sureste). Se puede apreciar la distribución.

a) mediante la distancia que separa la diagonal y la curva, la cual es una impresión gráfica

b) por el coeficiente gini que, entre el 0 y el 1, está más cercano al último. Así es que tenemos dos indicadores para la descripción de esta distribución, ambos muy simples de entender y apreciar.

Pasamos ahora a otra técnica muy usada en las ciencias sociales para medir la relación que puede existir entre dos conjuntos de datos. Es una manera de conocer la correlación o no correlación de datos. Se llama el Coeficiente Spearman de Correlación de Rangos (CSCR).

Para nuestro ejemplo se podrían usar dos conjuntos cualesquiera de datos para ver si entre sí tienen una correlación positiva o negativa. En otras palabras, cómo un grupo de factores influye sobre otros factores, si dicha influencia existe. Usaremos las cifras de siete naciones de América Latina para nuestro ejemplo. El ejercicio será aprender la aplicación de la fórmula:

$$r = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n^3 - n}$$

que se lee: "el rango es igual a uno menos seis veces la suma de los d_i al cuadrado partido por n al cubo menos n, donde d_i es la diferencia de rango y n es el número de casos examinados". Reconocemos de inmediato que habrá lectores que rehuyan de algo tan "complicado" como aparenta ser esta ingeniosa formulita, pero les rogamos tener algo de curiosidad y seguir adelante.

TABLA 2.0

Datos originales

Nación	% de la población alfabetizada	Ingresos per cápita
Brasil	61%	341
México	76	632
Argentina	93	914
Venezuela	63	837
Colombia	73	358
Perú	61	363
Chile	84	613

En esta tabla tenemos dos columnas de datos que pueden o no tener alguna correlación entre sí, y si la tienen, ésta puede servir para describir de manera más precisa los fenómenos bajo observación. A primera vista se puede decir "parece que hay correlación", pero no con certeza. Por ejemplo, un "todólogo" puede decir que es evidente que una baja tasa de alfabetismo implica una producción baja y por tanto los ingresos serían bajos. Pero el investigador serio no puede aceptar este argumento, y por tanto tiene que examinar los datos con mucho más precisión antes de postular una conclusión. La ventaja de trabajar organizadamente es que después de hacer el análisis, hay una base científica (al igual que el matemático o el físico), sobre la cual se puede discutir cualquier hipótesis. El "todólogo" solamente puede apreciar un fenómeno y emitir juicios. En este caso específico, si se comprueba que hay correlación hay que ver si es positiva o negativa, y qué grado de significancia tiene.

En el caso a mano, como se indica en la tabla 2.0, hemos escogido los datos de siete naciones latinoamericanas para nuestro ejemplo. Están ordenados arbitrariamente, tal como se podrían encontrar en un texto. Las cifras demuestran que Brasil tiene un 61% de su población alfabetizada y también tiene ingresos de 341 dólares por persona al año. De este modo están descritas otras seis naciones del área. Pero lo que queremos saber con algún grado de seguridad es si hay una correlación entre una cosa y la otra. Determinar la existencia de una correlación facilitaría cualquier descripción del área y al mismo tiempo se establecerían ciertos parámetros para planificadores y analistas.

Primer Paso

El primer paso es reorganizar los datos de una manera más coherente. Por ejemplo, vamos a hacer la reorganización de mayor a menor grado en cada variante: el grado de alfabetismo más alto recibirá el rango uno y el ingreso per cápita más alto será el número uno en esa columna.

TABLA 2.1

Datos originales		Datos reorganizados por rango	
61	341	93 (1)	914 (1)
76	632	84 (2)	837 (2)
93	914	76 (3)	632 (3)
63	837	73 (4)	613 (4)
73	358	63 (5)	363 (5)
61	363	61 (6.5)	358 (6)
84	613	61 (6.5)	341 (7)

La reorganización ha producido un cuadro artificial que nos permite colocar a cada valor en su lugar.

El caso especial en las últimas dos cifras en la columna de la alfabetización es la solución que se utiliza cuando hay dos valores iguales: se le asigna a cada uno de los valores el promedio de los rangos que le corresponden. Si hay n iguales, entonces se le asigna a cada uno un rango igual al promedio de los rangos que les corresponden.

Segundo Paso

Ya sabemos los rangos de los valores que estamos analizando, podemos colocar los datos en su posición original, pero ahora utilizando sus rangos para representar los verdaderos valores.

TABLA 2.2

Nación	Alfabetización Rangos	Ingresos per cápita Rangos
Brasil	6.5	7
México	3	3
Argentina	1	1
Venezuela	5	2
Colombia	4	6
Perú	6.5	5
Chile	2	4

Tercer Paso

El tercer paso es restar los rangos respectivos de la derecha y de la izquierda para obtener "d_i". El resultado es un número real.

TABLA 2.3

Nación	Alfabetización Rangos	Ingresos per cápita Rangos	d
Brasil	6.5	7	-0.5
México	3	3	0
Argentina	1	1	0
Venezuela	5	2	3
Colombia	4	6	-2
Perú	6.5	5	1.5
Chile	2	4	-2

Cuarto Paso

Este paso es sencillamente elevar al cuadrado (multiplicar por sí mismo) las cifras de la columna d y sumar los resultados.

d	d^2
$(-0.5)^2$	0.25
$(0)^2$	0
$(0)^2$	0
3^2	9
$(-2)^2$	4
$(1.5)^2$	2.25
$(-2)^2$	4
$\sum_{i=1}^{n} d_i^2 =$	19.50

Examinando las tablas desde 2.0 hasta 2.4, podemos ver que tenemos los valores necesarios para poder aplicar nuestra fórmula. En la primera tabla notamos que hay siete casos, n = 7. En las tablas subsiguientes hasta llegar a la 2.4 obtuvimos los datos necesarios para calcular la suma de los d_i^2 que es 19.50. Sustituyendo en la fórmula,

$$r = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n - n}, \text{ se obtiene } r = 1 - \frac{6\ (19.50)}{7^3 - 7} = 1 - \frac{117}{343 - 7} =$$

$$= 1 - \frac{117}{336} = 1 - 0.34 = 0.66.$$

El Coeficiente Spearman de Correlación de Rangos es 0.66 para el caso nuestro.

La pregunta que viene al caso ahora es: ¿Para qué sirve este número? Vamos a explicar su empleo.

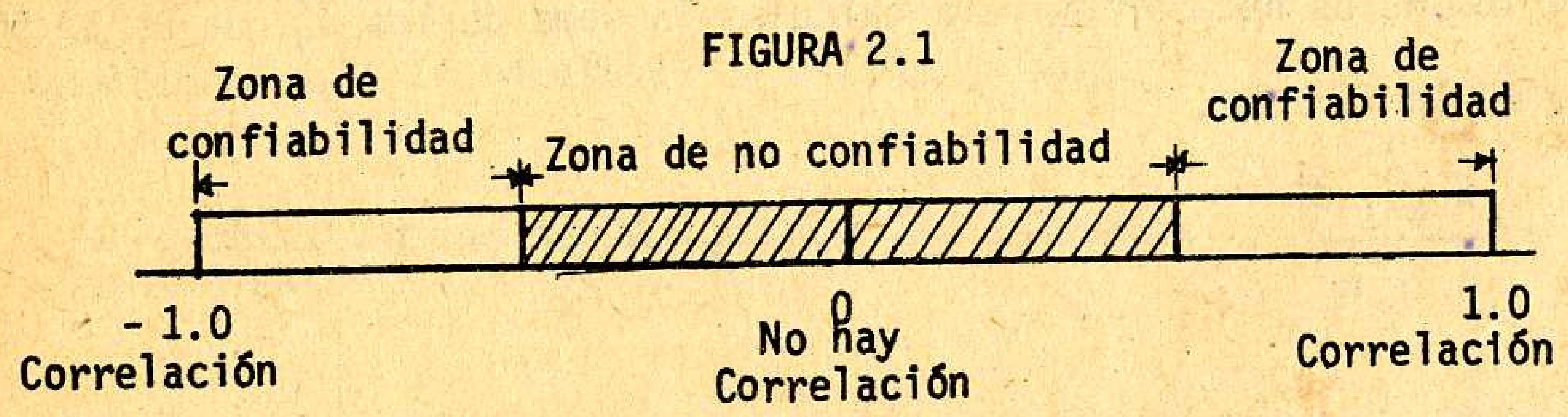
El resultado es una correlación positiva (0.66). ¿Qué quiere decir esto? Sencillamente, que cuando aumenta o disminuye una de las variables, la otra hace lo mismo. Si fuera por ejemplo una correlación negativa, entonces cuando una de las variables aumentara la otra disminuiría y viceversa. En nuestro caso particular, hemos logrado un indicio relativamente firme de que si se mejora el porcentaje de alfabetizados en una nación, también eventualmente existirá una mejoría en los ingresos per cápita. Otro modo de ver la misma situación es la siguiente: si conocemos el valor de una de las variables, entonces casi conocemos el de la otra.

Bueno, de eso hay que hablar un poco. Recordamos que el valor de r puede ser un número real que varía de -1 hasta +1. En general, la mayoría de los investigadores interpretan los valores de r de la manera siguiente:

a) Si el valor de r está entre -0.5 y +0.5, entonces la correlación no está muy bien definida.

b) Cuando el valor de r está entre -1 y -0.5 ó entre +0.5 y 1, entonces se afirma que hay correlación negativa o positiva respectivamente. La confiabilidad de la correlación existente aumenta cuando r se acerca a -1 ó a +1.

FIGURA 2.1



En el caso particular nuestro, tratándose solamente de siete naciones de América Latina, r = 0.66 es significativo pero no absolutamente. La razón de esta anomalía se encuentra en los casos particulares de Venezuela y Colombia. En Venezuela hay un elevado ingreso per cápita debido a su riqueza petrolera, pero al mismo tiempo hay un considerable analfabetismo. En Colombia hay una elevada tasa de alfabetismo, pero los ingresos per cápita son relativamente bajos.

Todo lo anterior presenta al investigador ciertos problemas, pues tendría que buscar otros datos antes de pontificar que alfabetismo y prosperidad van mano a mano.

Para mostrar el proceso que hemos descrito, he aquí otros datos que presentan la metodología en un solo cuadro.

TABLA 2.5

Nación	Consumo de Calorías per día per cápita	Consumo de Energía per día per cápita	Rango de datos		d	d^2
ARG	3060	1727	1	2	1-2=-1	1
BOL	1900	210	10	10	10-10=0	0
BRA	2620	535	3	7	3-7=-4	16
CHI	2670	1516	2	3	2-3=-1	1
COL	2200	595	7	6	7-6= 1	1
RD	2130	265	8	9	8-9=-1	1
ECUA	2010	296	9	8	9-8= 1	1
MEX	2580	1276	4	4	4-4= 0	0
PERU	2320	637	6	5	6-5= 1	1
VEN	2430	2433	5	1	5-1= 4	16
n = 10					$\sum_{i=1}^{n} d^2 =$	38

Utilizando de nuevo:

$$r = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n^3 - n}$$

$$= 1 - \frac{6\ (38)}{10^3 - 10}$$

$$= 1 - \frac{228}{1{,}000 - 10}$$

$$= 1 - \frac{228}{990}$$

$$= 1 - .23$$

$$= .77$$

Consultando la Figura 2.1 podemos ver que el resultado de .77 es significativo y positivo. Es decir, que podemos confiar bastante al decir que si una nación consume mucha energía, entonces debe tener también un alto consumo de calorías, y el recíproco también debe ser válido. Luego, la proposición inversa también es válida.

En este ejemplo, también se notan dos casos anómalos: Venezuela y Brasil. Para poder entender estos casos, el investigador tiene que ampliar sus parámetros de observación un poco más. En el caso de Venezuela, como es sabido, hay una enorme producción de petróleo, la cual exige un gasto de energía muy elevado para su extracción y procesamiento. Sin embargo, la riqueza producida por esa explotación ha creado el estado más urbanizado en América, pero sin un correspondiente aumento en la producción alimenticia por los hombres que han quedado en el campo. Por lo tanto, Venezuela es una nación importadora de alimentos, la mayoría del cual es consumido por un bajo porcentaje de la población. El caso de Brasil es lo contrario. Brasil, siguiendo su tradicional orientación hacia la exportación, envía una cantidad apreciable de comestibles de alto valor de calorías a sus mercados de ultramar. Por otro lado, el consumo de la energía en la nación está concentrado en la región sur (Sao Paulo y Río de Janeiro); pero en el sector industrial, y no para el consumo individual.

Como seguramente han apreciado los lectores, un solo instrumento no puede resolver todos los problemas del investigador. En cambio, varios de estos sencillos instrumentos pueden aumentar en gran parte el valor científico de su trabajo.

ANEXO I

El cálculo de los Coeficientes Spearman de Correlación de Rangos para 24 países de América Latina:

	n_1	n_2	Rangos			
Nación	Alfabetización	Ingresos	Alfabetización	Ingresos	d	(d^2)
BRASIL	61	341	17.5	14	17.5 - 14 = 3.5	12.25
MEXICO	76	632	8.5	6	8.5 - 6 = 2.5	6.25
ARGENT	93	914	1	2	1 - 2 = -1	1
VENEZ	63	837	15	3	15 - 3 = 12	144
COLOM	73	358	12.5	13	12.5 - 13 = -.5	.25
PERU	61	363	17.5	12	17.5 - 12 = 5.5	30.25
CHILE	84	613	3.5	8	3.5 - 8 =-4.5	20.25
C RICA	84	501	3.5	10	3.5 - 10 =-6.5	42.25
SALVAD	49	271	20	18	20 - 18 = 2	4
GUATEM	29	338	23	15	23 - 15 = 8	64
HOND	45	249	21	20	21 - 20 = 1	1
NICAR	50	387	19	11	19 - 11 = 8	64
PANAMA	73	629	12.5	7	12.5 - 7 = 5.5	30.25
CUBA	78	290	7	17	7 - 17 =-10	100
REP DOM	64	270	15	19	15 - 19 =-4	16
HAITI	10	70	24	24	24 - 24 = 0	0
JAMAICA	82	545	5	9	5 - 9 =-4	16
P RICO	81	1650	6	1	6 - 1 = 5	25
TRI-TOB	74	700	10.5	5	10.5 - 5 = 5.5	30.25
BOLIVIA	40	180	22	23	22 - 23 =-1	1
ECUADOR	61	247	14	21	14 - 21 =-7	49
GUYANA	76	300	8.5	16	8.5 - 16 =-7.5	56.25
PARAG	74	230	10.5	22	10.5 - 22 =-11.5	132.25
URUGUAY	91	773	2	4	2 - 4 =-2	4
n = 24						$\sum_{n=1}^{n} d^2 = 849.5$

ANEXO II

A continuación se muestra lo que ocurre cuando hay varios datos con el mismo valor; se han utilizado los datos del Anexo I:

Número	Dato	Rango
1	93	1
2	91	2
3	84	3.5
4	84	3.5
5	82	5
6	81	6
7	78	7
8	76	8.5
9	76	8.5
10	74	10.5
11	74	10.5
12	73	12.5
13	73	12.5
14	67	14
15	64	15
16	63	16
17	61	17.5
18	61	17.5
19	50	19
20	49	20
21	45	21
22	40	22
23	29	23
24	10	24

Vamos a utilizar un caso ficticio para mostrar qué sucede cuando hay tres datos con el mismo valor. En este caso lo que debe hacerse es sumar los rangos iguales: $2 + 3 + 4$ y dividir por el número de rangos sumados, esto es: $9 \div 3 = 3$. Entonces, a cada uno de los datos se le asigna el rango 3.

Número	Dato	Rango
1	5	1
2	3	3
3	3	3
4	3	3
5	1	5

Aplicamos la fórmula: $r = 1 - \frac{\sum_{i=1}^{n} d_i^2}{n^3 - n}$

$$= 1 - \frac{6\ (849.5)}{(24)^3 - 24}$$

$$= 1 - \frac{5097}{13{,}8000}$$

$$= 1 - .369$$

$$= .631$$

El resultado es algo menos significativo que nuestro ejemplo, pero debido al tamaño de la muestra es más confiable. Se puede notar la existencia de varios casos anómalos que bajan la relación positiva: Venezuela, Cuba, Guyana y Paraguay, por ejemplo.

Fuentes de datos:

Anuario Estadístico de las Naciones Unidas (UNSYB), 1971. New York: 1972.

Anuario Demográfico de las Naciones Unidas (UNDYB), 1970.

Population Reference Bureau, 1971.

J. P. Cole. Latin American: An Economic and Social Analysis. Totowa, N.J.: 1975.