ESTADISTICA ESPAÑOLA
núm. 100, 1983, págs. 105 a 110

Indices de concentración basados en la curva de Lorenz

JOSE MARIA ALBA RIESCO
Dpto. de Estadística y Econometría
(Universidad de Sevilla)

RESUMEN

Se realiza un estudio comparativo de diversas expresiones que aparecen en la literatura habitual, destinadas a expresar numéricamente la concentración de una distribución. La comparación y explicación de las distintas fórmulas se realiza analizando los principios en que se basan cada una de ellas, y al final se comparan en un ejemplo numérico. Las expresiones analizadas son las utilizadas en Alcaide (1973), Pulido (1976) y la inicialmente propuesta por Gini.

Palabras clave: Medidas de concentración, curva de Lorenz, índice de Gini.

1. INDICE DE CONCENTRACION DE GINI

Considérese una distribución de frecuencias con los datos agrupados en k intervalos $(l_{i-1}, l_i']$ con marcas de clase x_i y frecuencia absoluta de cada intervalo f_i. Sea N el número total de datos, y representamos por

$$y_i' = \frac{1}{N}\sum_{j=1}^{i} f_j$$

la frecuencia acumulada relativa, correspondiente al intervalo $(l_{i-1}, l_i]$, es decir, la proporción de datos que son menores o iguales que l_i.

Si $x_i f_i$ representa el total correspondiente a cada intervalo $(l_{i-1}, l_i]$, designemos por

$$\alpha_i = \frac{x_i f_i}{N\bar{Y}}$$

la proporción en que contribuye el intervalo $(l_{i-1}, l_i]$ al total de la distribución.

La proporción en que contribuyen los datos que son menores o iguales que l_i al total de la distribución es

$$q_i = \sum_{j=1}^{i} \alpha_j$$

Obviamente $v'_0 = 0$, $v'_k = 1$ y $q_0 = 0$, $q_k = 1$. La poligonal que une los puntos de coordenadas (v'_i, q_i), $i = 0, 1, 2, \ldots, k$ es la curva de concentración de Lorenz y el doble del área determinada por esta poligonal y la recta de equidistribución es, por definición, el índice de concentración de Gini.

Así pues, para la determinación del índice de concentración de Gini es suficiente con determinar este área. Es fácil de obtener, entonces, la expresión del índice de Gini

$$I_G = \sum_{i=1}^{k} \alpha_i (v'_i + v'_{i-1} - 1) \qquad [1]$$

2. INDICE DE CONCENTRACION DE ALCAIDE

He dado esta denominación al índice que se obtiene al aplicar el razonamiento expuesto en Alcaide (1973) a una distribución de frecuencias de datos agrupados en intervalos de clase. Con la notación ya introducida se obtiene

$$I_A = 1 - \frac{1 - \sum_{i=1}^{k} \alpha_i v'_i}{1 - \sum_{i=1}^{k} f'_i v'_i} \qquad [2]$$

siendo $f'_i = \dfrac{f_i}{N}$ la frecuencia relativa del intervalo $(l_{i-1}, l_i]$.

3. INDICE DE CONCENTRACION DE PULIDO

He dado esta denominación al índice que se obtiene cuando se aplica el método expuesto en Pulido (1976) a una distribución de frecuencias con datos agrupados en intervalos de clase.

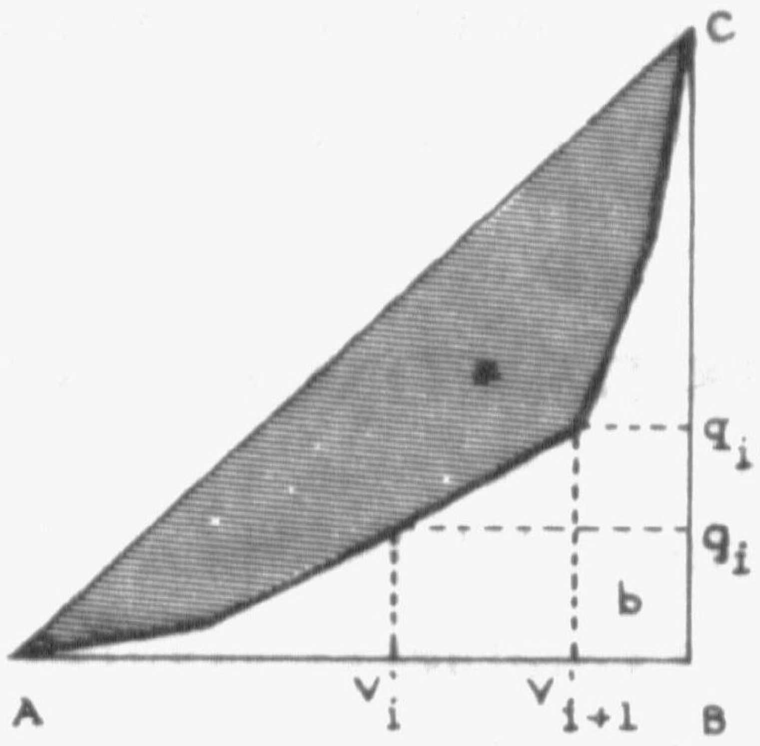
El resultado que se obtiene es

$$I_p = 1 - \frac{k - \sum_{i=1}^{k} i\alpha_i}{k - \sum_{i=1}^{k} if_i} \qquad [3]$$

4. JUSTIFICACION DE LAS DISCREPANCIAS

El origen de las discrepancias entre estos tres índices de concentración está en el principio que se ha seguido para su determinación, principios que son expuestos a continuación.

Principio de Gini. Ya se ha indicado que I_G es el doble del área determinada por la curva de Lorenz y la recta de equidistribución, o bien el cociente entre este área y la determinada por los puntos A, B, C.

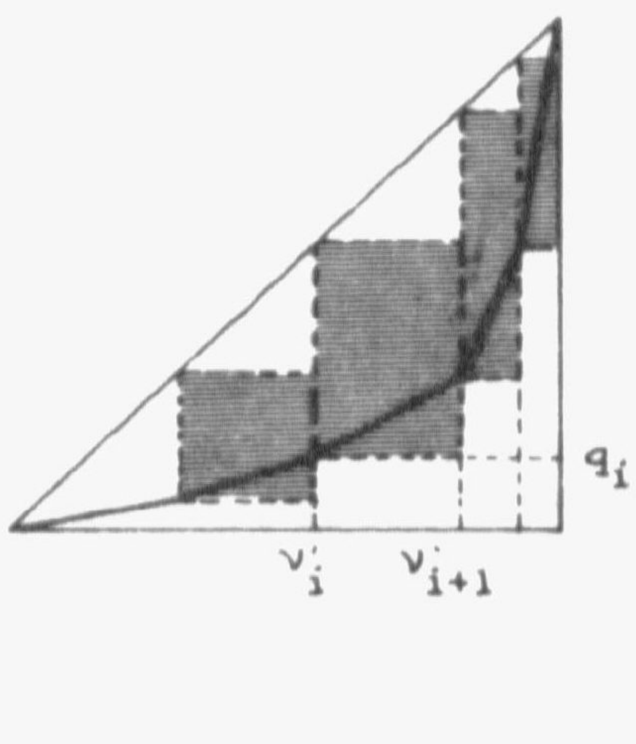


$$I_G = 2a = 2\left(\frac{1}{2} - b\right)$$

b = área determinada por A, B, C, y curva de Lorenz =

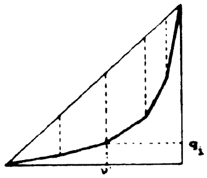
$$= \sum_{i=0}^{k-1} \frac{1}{2}(v'_{i+1} - v'_i)(q_i + q_{i+1})$$

Principio de Alcaide. El índice de concentración es el cociente entre la suma de las áreas de los rectángulos de base $(v'_{i+1} - v'_i)$ y altura $(\theta'_i - q_i)$, para $i = 1, 2, \ldots, k - 1$, y la suma de las áreas de los rectángulos de base $(v'_{i+1} - v'_i)$ y altura v'_i, para $i = 1, 2, \ldots$ $k - 1$.



$$I_A = \frac{\sum_{i=1}^{k-1} (v'_{i+1} - v'_i)(v'_i - q_i)}{\sum_{i=1}^{k-1} (v'_{i+1} - v'_i)v'_i}$$

Principio de Pulido. El índice de concentración viene determinado por el cociente entre las suma de las distancias, medidas en sentido vertical, desde el punto (v'_i, q_i) de la curva de Lorenz a la recta de equidistribución para $i = 1, 2, \ldots, k - 1$, y la suma de estas distancias para el caso hipotético de máxima concentración, es decir



$$I_P = \frac{\sum_{i=1}^{k-1} (v'_i - q_i)}{\sum_{i=1}^{k-1} v'_i}$$

Se ve, pues, que la diferencia entre estos índices de concentración estriba en el principio o definición que se adopta. Todos son índices de concentración en el sentido de que verifican las siguientes propiedades:

a) Son coeficientes independientes de las magnitudes que se consideren.

b) Su campo de variación es el intervalo [0, 1], correspondiendo el valor mínimo 0 al caso en que no hay concentración (la curva de Lorenz coincide con la curva de equidistribución) y el valor máximo, 1, al caso extremo de máxima concentración en que la curva de Lorenz pasa por el punto (1, 0). El índice de Pulido presenta la gran ventaja de ser muy sencillo de calcular, mientras que los índices de Alcaide y de Gini requieren un cálculo más laborioso.

Esta ventaja no es tal hoy día gracias a las facilidades de cálculo que proporciona cualquiera de las pequeñas máquinas de calcular existentes en la actualidad.

5. COMPARACION ENTRE LOS DISTINTOS INDICES DE CONCENTRACION

Dado que estos índices son obtenidos a partir de principios o definiciones diferentes, cuando los datos están agrupados en intervalos no es posible establecer comparaciones entre ellos. Sin embargo, en el caso particular de que la agrupación de los datos se haya llevado a cabo de modo que todos los intervalos tengan la misma frecuencia, los índices sí son comparables. Así, si $f_i = \frac{N}{k}$, las fórmulas [1],[2] y [3] anteriores conducen a las siguientes expresiones:

$$I_G = \frac{1}{k} [2 \sum_{i=1}^{k} i\alpha_i - (k + 1)] \qquad [4]$$

$$I_A = I_P = \frac{1}{k - 1} [2 \sum_{i=1}^{k} i\alpha_i - (k + 1)] \qquad [5]$$

es decir, coinciden los valores de los índices de Alcaide y de Pulido, pero su valor es superior al índice de Gini:

$$I_G < I_A = I_P = \frac{k}{k-1} I_G.$$

Es decir, cuando los datos están agrupados en intervalos de igual frecuencia, el índice de Alcaide y el índice de Pulido coinciden y dan un valor superior al índice de Gini.

Esta situación se presenta cuando los datos están sin agrupar (k = N), o bien, cuando están agrupados por cuantiles (k = 10 (deciles); k = 4 (cuartiles), etc.).

La fórmula [4] para k = 10, es decir, para datos agrupados por deciles coincide con la dada por Uriel (1974).

6. COMPROBACIONES NUMERICAS

La disparidad de las discrepancias entre estos tres índices de concentración se pone de manifiesto claramente cuando se calculan los tres para una misma distribución.

Considerando las distribuciones de las rentas salariales en los años 1971, 1972 y 1973, dadas por el INE en el cuadro A-II-14 de La Renta Nacional en 1975 y su distribución, y que se adjunta en el anexo, los valores de los índices de concentración calculados por las fórmulas [1], [2] y [3] son:

INDICE DE CONCENTRACION DE

Dato de	Gini	Alcaide	Pulido
1971	0.3280	0.3500	0.2917
1972	0.3263	0.3489	0.3259
1973	0.3297	0.3671	0.3803

7. ANEXO

ESTRUCTURA DE LAS RENTAS SALARIALES (1)

INGRESOS MEDIOS MENSUALES, INCLUIDA LA AYUDA FAMILIAR	1971		1972		1973	
	Porcentaje de trabajadores	Porcentaje de ingresos	Porcentaje de trabajadores	Porcentaje de ingresos	Porcentaje de trabajadores	Porcentaje de ingresos
Menos de 3.000 pesetas	5,5	1,3	3,9	0,8	3,0	0,5
De 3.001 a 4.000 pesetas	6,2	2,4	4,5	1,5	3,6	1,0
De 4.001 a 5.000 pesetas	11,3	5,6	7,8	3,4	4,6	1,7
De 5.001 a 6.000 pesetas	11,2	6,8	9,5	5,0	6,7	3,0
De 6.001 a 7.000 pesetas	10,3	7,4	9,6	5,9	7,9	4,0
De 7.001 a 8.000 pesetas	9,1	7,4	8,9	6,3	8,1	4,9
De 8.001 a 9.000 pesetas	7,9	7,3	7,8	6,3	7,7	5,3
De 9.001 a 10.000 pesetas	6,8	7,1	6,8	6,1	6,6	5,1
De 10.000 a 11.000 pesetas	5,6	6,4	6,1	6,1	6,3	5,3
De 11.001 a 12.000 pesetas	5,0	6,3	5,6	6,1	5,7	5,3
De 12.001 a 14.000 pesetas	6,8	9,6	8,1	9,9	9,0	9,4
De 14.001 a 16.000 pesetas	4,5	7,9	6,1	8,7	7,5	9,1
De 16.000 a 18.000 pesetas	3,1	5,8	4,5	7,3	6,3	8,7
De 18.001 a 20.000 pesetas	2,1	4,4	3,4	6,0	4,7	7,7
Más de 20.000 pesetas	4,6	14,8	7,4	20,6	12,3	29,0
Totales	100,0	100,0	100,0	100,0	100,0	100,0
Indice concentración de GINI	0,3243		0,3263		0,3281	

(1) Se refiere a los sectores comprendidos en la estadística de salarios y que se analizan en los cuadros.
FUENTE: INE. Encuesta de salarios.

8. REFERENCIAS

ALCAIDE, A.: *Estadística Económica*. Ed. Saeta, Madrid, 1976.

PULIDO SAN ROMÁN, A.: *Estadísticas y Técnicas de Investigación Social*. Ed. Pirámide, Madrid, 1976.

URIEL, E.: «Comparación entre diversas medidas de desigualdad». *Estadística Española*, núm. 64-65, 39-69.

SUMMARY

In this paper we study several measures of concentration of a distribution. Thus, we discusses the measures of concentration used by Alcaide (1973), Pulido (1976) and Gini. Finally, by analysing some economic data, we compare such measures.

Key words: Measures of Concentration; Lorenz's curve; Gini's Index.

AMS 1970. Subjet classification: 62-00, 62P99.