Problemas de agregación en el coeficiente de concentración de Gini

Basulto Santos, Jesús

Universidad de Sevilla.

e-mail: basulto@us.es

Romero García, José Enrique.

Universidad de Sevilla.

e-mail: romerogje@us.es

Resumen

Hemos realizado un análisis de los problemas que pueden producirse al calcular la concentración de una característica cuantitativa cuando en lugar de trabajar con los datos desagregados observados se realiza la agregación de dichos datos.

Partiremos del artículo de Gini, de 1914, en donde define su *coeficiente de concentración* como una media aritmética ponderada de las denominadas *desigualdades relativas*. La formula que propone Gini la limita a datos desagregados y, posteriormente, la modifica tanto para datos agregados en frecuencias como para datos agregados en intervalos.

Palabras clave: Lorenz, dispersión, concentración, Gini.

Área Temática: Metodología

1. Introducción.

El cálculo de la concentración para valores desagregados ha sido realizado aplicando generalmente de forma correcta la fórmula original de Gini; pero no ha ocurrido igual con el cálculo de la concentración para valores agregados, para los que, en algunas ocasiones, se han aplicado fórmulas incorrectas. En el presente trabajo, hemos intentado esclarecer distintas expresiones para el cálculo de la concentración y su aplicabilidad según las circunstancias. Previamente a este trabajo ha sido realizado un análisis profundo del artículo original de Gini "Sulla misura della concentrazione e della variabilità dei caractteri" (1914). Este trabajo fue consecuencias de los trabajos: *Indici di Concentrazione e di dipendenza* de 1910 y Variabilità e Mutabilità de 1912. A partir de aquí, recogemos la fórmula original de Gini para datos desagregados en frecuencias unitarias en la sección 2; en la sección 3 recogemos la fórmula de Gini para datos agregados en frecuencias, dando una nueva expresión de esta fórmula en la sección 4; en la sección 5 recogemos la fórmula de Gini para datos agregados en intervalos y, por último, finalizamos con las principales conclusiones.

2. Datos desagregados.

Nos proponemos en esta sección dar una justificación, basada en un cuidadoso estudio del artículo histórico de Gini de 1914, del denominado habitualmente *índice de Gini* para el cálculo de la concentración para valores desagregados, cuya expresión suele recogerse en los libros de texto como $R = \dfrac{\sum_{i=1}^{n-1}(P_i - Q_i)}{\sum_{i=1}^{n-1} P_i}$.

El objetivo perseguido por Gini en su trabajo de 1914 es, en sus propias palabras: "*proponer una medida de concentración que sea independiente de la curva de distribución del carácter*". En su artículo, Gini considera que disponemos de n valores ordenados $\{a_i, i=1,..,n\}$ de una variable estadística cuantitativa no negativa, es decir, a_1=..... = a_n. Así, si definimos las variables $A_i = \sum_{k=1}^{i} a_k$, cantidad acumulada del carácter hasta el lugar i-ésimo, las variables $Q_i = \dfrac{A_i}{A_n} = \dfrac{\sum_{k=1}^{i} a_k}{\sum_{k=1}^{n} a_k}$, que constituyen la proporción que representa la cantidad acumulada del carácter hasta el lugar i, A_i, respecto a la cantidad total presente del carácter, y las variables $P_i = \frac{i}{n}$, que denotan la proporción de observaciones hasta el lugar i respecto al número de observaciones totales, puede verificarse fácilmente que

$$Q_i \leq P_i .$$

Gini afirma:

"la concentración del carácter será tanto más fuerte cuanto más fuerte sean las n-1 desigualdades $Q_i < P_i$, i=1,..,n-1". (4)

La desigualdad (4) es tanto más fuerte, en valor absoluto, cuanto mayor sea la diferencia P_i - Q_i; y , dado que el valor máximo de Q_i es P_i, tanto más fuerte en valor relativo cuanto más alta es la razón

$$R_i = \frac{P_i - Q_i}{P_i} .$$

Analicemos seguidamente cuál es el significado de $R_i P_i = P_i - Q_i$.

Veamos que es la proporción, respecto a la intensidad total de todo el colectivo, que ha de transferirse al colectivo de los P_i casos con intensidad más baja, para eliminar la desigualdad presente en los mismos respecto al conjunto del colectivo.

En efecto,

$$Q_i^* = \frac{A_i^*}{A_n^*} = \frac{\sum_{k=1}^{i} a_k + (P_i - Q_i)\sum_{k=1}^{n} a_k}{\sum_{k=1}^{n} a_k} =$$
$$= Q_i + (P_i - Q_i) = P_i .$$

En el caso de equidistribución, la proporción acumulada del carácter, Q_i , sería igual a la proporción acumulada de individuos, P_i; luego el cociente $R_i = \frac{P_i - Q_i}{P_i}$ tendría que ser cero. Por el contrario, en el caso de máxima desigualdad, la proporción acumulada del carácter, Q_i, sería igual a cero, luego $P_i - Q_i = P_i$ y el cociente $R_i = \frac{P_i - Q_i}{P_i}$ tendría que ser uno. Así pues, el valor de R_i varía entre 0, en caso de equidistribución, y uno, en caso de concentración perfecta; y, por ello, puede considerarse un indicador de la concentración del carácter que se presenta entre los P_i casos que tienen intensidad más baja en el carácter en estudio.

Dado que ese indicador de la concentración del carácter afecta a la proporción P_i de casos, es por lo que como medida de la concentración global del carácter podremos asumir la media ponderada de los n-1 valores de R_i, donde cada uno de los R_i entra con un peso proporcional al valor P_i.

Tal medida será

$$R = \frac{\sum_{i=1}^{n-1} R_i P_i}{\sum_{i=1}^{n-1} P_i} = \frac{\sum_{i=1}^{n-1} \frac{P_i - Q_i}{P_i} P_i}{\sum_{i=1}^{n-1} P_i} = \frac{\sum_{i=1}^{n-1} (P_i - Q_i)}{\sum_{i=1}^{n-1} P_i},$$

es decir:

$$R = \frac{\sum_{i=1}^{n-1} (P_i - Q_i)}{\sum_{i=1}^{n-1} P_i} \ . \tag{11}$$

Esa medida, que en los libros de texto habituales aparece como índice de Gini, es la que Gini denomina *Razón de Concentración*.

Obsérvese, que para concentración perfecta, Q_i=0, i=1,..,n-1, con lo que $\sum_{i=1}^{n-1} (P_i - Q_i) = \sum_{i=1}^{n-1} P_i$, luego R=1. Para equidistribución, P_i=Q_i, con lo que R=0.

Obsérvese también, que a medida que las desigualdades $Q_i < P_i$ son más fuertes, entonces R aumenta.

La razón de concentración es por lo tanto, un indicador de concentración que verifica las condiciones siguientes: crece al crecer la concentración, toma como valor más alto uno en el caso de máxima concentración y toma su valor más pequeño, 0, en el caso de mínima concentración.

Teniendo presente que

$$\sum_{i=1}^{n-1} P_i = \sum_{i=1}^{n-1} \frac{i}{n} = \frac{1}{n}\sum_{i=1}^{n-1} i = \frac{1}{n}\frac{(n-1)\cdot n}{2} = \frac{(n-1)}{2},$$

obtenemos la siguiente formulación para *R*, más operativa que la original:

$$R = \frac{2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n} - 1 \qquad \text{(12 bis).}$$

3. Agregación en frecuencias.

Si el cálculo de la concentración para valores desagregados ha sido realizado aplicando generalmente de forma correcta la fórmula original de Gini, no ha ocurrido igual, como indicamos en la introducción, con el cálculo de la concentración para valores agregados. En muchos manuales de estadística, se utiliza la fórmula

$$R^* = \frac{\sum_{i=1}^{s-1}(P_i - Q_i)}{\sum_{i=1}^{s-1} P_i}\ , \qquad \text{(I)}$$

a la que también se le asigna el nombre de índice de Gini para datos agregados en frecuencias; y, como veremos, está fórmula nunca fue usada por Gini, y además puede presentar serías deficiencias.

En lo que sigue, obtendremos la expresión de la razón de concentración, *R*, para el caso que los *n* valores estén agrupados en frecuencias absolutas; es decir, que de los *n* datos solo haya *s* valores diferentes x_l, $l=1,..,s$, y cada valor x_l se repite n_l veces (Gini los llama f_l).

Para obtener la fórmula, Gini define los valores N_l, como el número total de individuos con valores menores o iguales a x_l, *y* N_{l-1} , como el número de individuos con valores menores a x_l, es decir menores o iguales a x_{l-1} (él los llama i_l e i_{l-1}); y por tanto, es claro que $N_{l-1} = N_l - n_l$.

Con lo cual, partiendo de (12 bis), se obtiene que

$$R = \frac{\sum_{l=1}^{s}(N_{l-1} + N_l - 1)\, x_l n_l}{(n-1)A_n} - 1\,. \qquad (13)$$

En muchos libros de texto, se utiliza la fórmula (11) de Gini con datos agregados en frecuencias, es decir, la fórmula (I); pero (I) no coincide con (13) en estos casos de datos agregados en frecuencias.

En efecto, consideremos el siguiente ejemplo numérico en el que se aprecia claramente los diferentes resultados que aportan ambas expresiones. Calculemos en primer lugar *R* mediante (13):

xi	ni	Ni	Si=xini	Ni-1+Ni -1	(Ni-1+Ni -1)xini
5	15	15	75	14	1050
80	90	105	7200	119	856800
	105		7275		857850
				R=	0.133822363

Se obtiene un valor para *R* de 0.13, indicando que hay poca concentración, cosa lógica pues hay pocos individuos con ingresos pequeños y muchos individuos con elevados ingresos.

Si ahora calculamos *R* usando la fórmula (I), se obtiene:

xi	ni	Ni	ti=xini	Ai	Pi	Qi
5	15	15	75	75	0.142857143	0.01030928
80	90	105	7200	7275	1	1
	105		7275			
				R*=	0.927835052	

Que indica concentración muy elevada, cosa que sabemos que no ocurre.

4. Otra formulación del índice R para datos agregados en frecuencias absolutas

Nos hemos propuesto una expresión alternativa que sea más operativa a la dada por Gini en (13). Para ello, partiremos de su fórmula original, $R = \dfrac{\sum_{i=1}^{n-1}(P_i - Q_i)}{\sum_{i=1}^{n-1} P_i}$.

Para cada uno de los s valores diferentes presentes en los n datos, x_i, consideramos las cantidades

$$L_i = \sum_{j=N_{i-1}+1}^{N_{i-1}+n_i} P_j = n_i \left[P_i - \frac{(n_i - 1)}{2N} \right]$$

$$M_i = \sum_{j=N_{i-1}+1}^{N_{i-1}+n_i} Q_j = n_i \left[Q_i - \frac{(n_i - 1)\, x_i}{2A_n} \right].$$

Puede comprobarse que

$$\sum_{i=1}^{n-1} (P_i - Q_i) = \sum_{i=1}^{s} (L_i - M_i)$$

$$\frac{n-1}{2} = \sum_{i=1}^{n-1} P_i = \sum_{i=1}^{s} L_i - 1,$$

con lo que

$$R = \frac{\sum_{i=1}^{n-1} (P_i - Q_i)}{\sum_{i=1}^{n-1} P_i} = \frac{\sum_{i=1}^{s} (L_i - M_i)}{\sum_{i=1}^{s-1} L_i - 1} =$$

$$= \frac{\sum_{i=1}^{s} L_i}{\sum_{i=1}^{s-1} L_i - 1} - \frac{\sum_{i=1}^{s} M_i}{\sum_{i=1}^{s-1} L_i - 1} =$$

$$= \frac{\sum_{i=1}^{s} L_i - 1}{\sum_{i=1}^{s-1} L_i - 1} + \frac{1}{\sum_{i=1}^{s-1} L_i - 1} - \frac{\sum_{i=1}^{s} M_i}{\sum_{i=1}^{s-1} L_i - 1} =$$

$$= 1 + \frac{1}{(n-1)/2} - \frac{\sum_{i=1}^{s} M_i}{(n-1)/2} =$$

$$= \frac{n + 1 - 2\sum_{i=1}^{s} M_i}{n-1};$$

es decir,

$$R = \frac{n+1-2\sum_{i=1}^{s} M_i}{n-1}, \qquad \text{(II)}$$

con $M_i = \eta_i \left[Q_i - \frac{(n_i - 1)x_i}{2A_n} \right]$.

Igualmente, hemos obtenido que una expresión de R a partir, exclusivamente, de los P_i, Q_i de los s valores diferentes presentes en los n datos originales viene dada por:

$$R = \frac{\sum_{i=1}^{s} n_i \left(\hat{P}_i - \hat{Q}_i\right)}{(N-1)/2} = \frac{2}{N-1}\left[\sum_{i=1}^{s} n_i (P_i - Q_i) + \frac{1}{2}\sum_{i=1}^{s} n_i (q_i - p_i)\right],$$

donde,

$$\hat{P}_i = P_i - \frac{p_i}{2}, \quad p_i = \frac{n_i}{N}$$
$$\hat{Q}_i = Q_i - \frac{q_i}{2}, \quad q_i = \frac{x_i n_i}{\sum_{j=1}^{s} x_j n_j}.$$

Si ahora calculamos la razón de concentración mediante (II):

xi	ni	Ni	ti=xini	Ai	Qi	Mi
5	15	15	75	75	0.010309278	0.08247423
80	90	105	7200	7275	1	45.9587629
	105		7275			46.0412371
				R=	0.133822363	

Es decir, hemos obtenido el mismo valor que el calculado usando la fórmula (13).

5. Agregación en intervalos.

Ahora, extenderemos la fórmula (12 bis) a situaciones donde los datos han sido agregados en r clases, con límites $I_k=[l_{k-1}, l_k]$, $k=1,..,r$.

Para ver en que se traduce, en este caso, la fórmula (12 bis), se definen las frecuencias acumuladas N_k y N_{k-1}, donde N_k es el número total de datos con valores menores o iguales que l_k , y N_{k-1} el número total de datos con valores inferiores o iguales a l_{k-1}. Igualmente define S_k como la cantidad total del carácter que está presente en la clase k-ésima.

La formula (12 bis) conduce, en esta ocasión, a esta otra fórmula aproximada:

$$R' = \frac{\sum_{k=1}^{r}\left[S_k\left(N_{k-1}+N_k-1\right)\right]}{\left(n-1\right)A_n} - 1 = \frac{\left[\sum_{k=1}^{r}S_k\left(N_{k-1}+N_k\right)\right]-nA_n}{\left(n-1\right)A_n}, \tag{15}$$

verificándose que:

$$R' < R. \tag{16}$$

Si consideramos que los n_k valores de la intensidad del carácter en la clase I_k siguen una progresión aritmética, lo cual es un caso particular de que los valores del carácter se distribuyan, dentro de cada clase, de forma uniforme; se tiene que, en el caso de agrupación en intervalos, y bajo el supuesto establecido de progresión aritmética, la expresión aproximada de R, es:

$$R'' = \frac{1}{\left(n-1\right)A_n}\left[\sum_{k=1}^{r}\left[S_k\left(N_{k-1}+N_k-1\right)+\frac{1}{6}\left(n_k^{\,2}-1\right)\left(l_k-l_{k-1}\right)\right]\right]-1, \tag{17}$$

Si construimos una curva de concentración con datos agregados en intervalos de clase, y a continuación se unen los puntos de cambio de la curva de concentración por segmentos; aproximando así la curva de concentración por una poligonal y luego calculamos la aproximación al área de concentración como el área existente entre la recta de equidistribución y la poligonal, veamos seguidamente la fórmula que se obtiene.

Puede comprobarse fácilmente, que el área de concentración aproximada, ACA, para el caso que la curva de concentración se aproxime por una poligonal, tiene la siguiente expresión:

$$ACA = \frac{\sum_{k=1}^{r}\left[S_k\left(N_{k-1}+N_k\right)\right]}{2nA_n} - \frac{1}{2} = \frac{\left[\sum_{k=1}^{r}S_k\left(N_{k-1}+N_k\right)\right]-nA_n}{2nA_n}. \tag{20}$$

Ahora, para calcular una medida aproximada de la concentración Gini dividió ese área entre el área del triangulo inferior que determina la gráfica de la curva de concentración, que como sabemos es ½, luego la medida aproximada de la concentración que obtiene Gini a partir de este método gráfico, que llamaremos índice de Gini Geométrico y notaremos por IG_G , es:

$$IG_G = \frac{ACA}{1/2} = 2ACM = \frac{\sum_{k=1}^{r}\left[S_k\left(N_{k-1}+N_k\right)\right]}{nA_n} - 1 = \frac{\left[\sum_{k=1}^{r} S_k\left(N_{k-1}+N_k\right)\right] - nA_n}{nA_n}.$$

Expresión que se relaciona con la fórmula de *R*’ hallada en (15) (a R’ le llamaremos IG_T, índice de Gini Teórico) que aproximaba a la razón de concentración *R*, pues

$$IG_G = \frac{n-1}{n}R'.$$

Si Gini hubiese dividido el área de concentración aproximada, ACA, por el área correspondiente a la situación de máxima concentración (que está determinada por el triángulo de vértices (0,0), ((n-1)/n,0) y (1,1)), que es $\frac{n-1}{2n}$, se hubiese obtenido, según (20):

$$\frac{ACA}{ACM} = \frac{ACA}{\frac{n-1}{2n}} = \frac{2n}{n-1}\left[\frac{\left[\sum_{k=1}^{r} S_k\left(N_{k-1}+N_k\right)\right] - nA_n}{2nA_n}\right] = \frac{\left[\sum_{k=1}^{r} S_k\left(N_{k-1}+N_k\right)\right] - nA_n}{(n-1)A_n},$$

expresión que coincide con el valor de *R*’ dado en (15).

En consecuencia, la aproximación *R*’ de *R*, procede de haber realizado la aproximación de la verdadera curva de concentración, que es la poligonal dada para datos desagregados en frecuencias unitarias (o no unitarias, pues coinciden en ambos casos) por la poligonal obtenida al agregar los datos en intervalos.

El propio Gini, recoge un ejemplo, realizado por Pietra, sobre la propiedad de la tierra en manos privadas en el estado de Victoria, sobre la influencia del número de clases en los valores de R’ y R”. Exponemos seguidamente los resultados obtenidos:

Número de clases	Valor de R’	Valor de R”
30	69.0%	69.1 %
25	68.9 %	69.2 %
20	68.8 %	69.2 %
15	68.4 %	69.1 %
10	67.8 %	69.2 %
8	66.5 %	69.1 %

6	65.8 %	69.6 %
5	64.6 %	69.3 %
4	58.1 %	71.1 %

Como se observa, la formula R” es una muy buena aproximación al valor verdadero de R, y es la que permanece más estable, independientemente del número de clases que se consideren.

Nosotros también hemos realizado pruebas para verificar la estabilidad de R”. A continuación mostramos algunos resultados.

datos	ni	Ni	(Ni-1 +Ni -1)xi ni	datos	ni	Ni	(Ni-1 +Ni -1)xi ni
61	2	2	122	102	10	189	374340
68	3	5	1224	103	5	194	196730
69	2	7	1518	104	4	198	162656
70	2	9	2100	105	6	204	252630
71	2	11	2698	106	1	205	43248
72	4	15	7200	107	2	207	87954
73	4	19	9636	108	1	208	44712
74	2	21	5772	109	5	213	228900
75	1	22	3150	110	4	217	188760
76	3	25	10488	111	4	221	194028
77	3	28	12012	113	2	223	100118
78	5	33	23400	114	1	224	50844
79	3	36	16116	115	3	227	155250
80	3	39	17760	118	3	230	161424
81	6	45	40338	119	1	231	54740
82	6	51	46740	120	3	234	167040
83	3	54	25896	121	1	235	56628
84	8	62	77280	122	1	236	57340
85	6	68	65790	123	1	237	58056
86	4	72	47816	125	1	238	59250
87	6	78	77778	126	2	240	120204
88	5	83	70400	127	1	241	60960
89	13	96	205946	128	1	242	61696
90	8	104	143280	129	2	244	125130
91	9	113	176904	130	1	245	63440
92	6	119	127512	132	3	248	194832
93	6	125	135594	136	600	848	89352000
94	7	132	168448	137	600	1448	188649000
95	5	137	127300	140	600	2048	293580000
96	6	143	160704	150	600	2648	422550000
97	5	148	140650	551	600	3248	1948887000
98	3	151	87612				
99	7	158	213444		3248		2949301278
100	9	167	291600				
101	12	179	418140		691827		

R= 0.31292326

Hemos obtenido que el verdadero valor de *R* es *R*=0.31

Si ahora agrupamos en 5 intervalos, obtenemos:

Intervalos	nk	Sk	Nk-1+Nk -1	Sk (Nk-1+Nk -1)	Pk	Qk
x<=79	36	2635	35	92225	0.01108374	0.00380876
79<x<=97	112	9955	183	1821765	0.0455665	0.01819819
97<x<=115	79	8222	374	3075028	0.06988916	0.03008267
115<x<=133	21	2615	474	1239510	0.07635468	0.03386251
133<x<=551	3000	668400	3495	2336058000	1	1
				2342286528	0.20289409	0.08595212
			R' =	0.042702043	R*=	0.5763695

Ik –Ik-1	Nk	Ak		Sk(Nk-1+Nk-1)+(nk2-1)*(Ik-Ik-1)/6
18	36	2635		96110
18	148	12590		1859394
18	227	20812		3093748
18	248	23427		1240830
418	3248	691827		2963057930
				2969348012
			R"=	0.32184735

Por lo que hemos obtenido los siguientes resultados: R'=0.04, R"=0.32, R*=0.57.

Si agrupamos en 10 intervalos, obtenemos:

	nk	Sk	Nk-1+Nk -1	Sk (Nk-1+Nk -1)	Pk	Qk
I<=70	9	604	8	4832	0.00277094	0.00087305
70-79	27	2031	44	89364	0.01108374	0.00380876
79-88	47	3955	118	466690	0.02555419	0.0095255
88-97	65	6000	230	1380000	0.0455665	0.01819819
97-106	57	5786	352	2036672	0.06311576	0.02656155
106-115	22	2436	431	1049916	0.06988916	0.03008267
115-124	10	1199	463	555137	0.07296798	0.03181576
124-133	11	1416	484	685344	0.07635468	0.03386251
	1800	247800	2295	568701000	0.63054187	0.39204454
	1200	420600	5295	2227077000	1	1
	3248			2802045955	0.99784483	0.54677253

R'= 0.24737047 R*= 0.45204653

Ik-Ik-1	Nk	Ak		Sk(Nk-1+Nk-)+(nk2-1)*(Ik-Ik-1)/6
9	9	604		4952
9	36	2635		90456
9	83	6590		470002
9	148	12590		1386336
9	205	18376		2041544
9	227	20812		1050640.5
9	237	22011		555285.5
9	248	23427		685524
9	2048	271227		573560999
409	3248	691827		2325236932
				2905082670

R”= 0.29323872

Así pues, hemos obtenidos, siendo *R*=0.31 el verdadero, los siguientes resultados:

Nº intervalos	R’	R”	R*
5 intervalos	0.042	0.32	0.57
10 intervalos	0.24	0.29	0.45

Que confirman la estabilidad de R”, la falta de precisión de R* y el alejamiento de R’ del verdadero R al disminuir el número de clases que hemos establecido.

7. Conclusiones .

En resumen, hemos obtenido las siguientes conclusiones sobre la aplicabilidad de las diversas fórmulas:

1. Para datos desagregados con frecuencias unitarias:

- $$R = \frac{\sum_{i=1}^{n-1} (P_i - Q_i)}{\sum_{i=1}^{n-1} P_i} \qquad (11)$$

- $$R = \frac{2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n} - 1 \quad \text{(12 bis)}$$

Tanto las fórmulas (11) como la (12 bis) son fórmulas exactas. La (12 bis) es más operativa.

2. Para datos agregados en frecuencias absolutas:

- $$R = \frac{\sum_{l=1}^{s}(N_{l-1} + N_l - 1)\, x_l n_l}{(n-1)A_n} - 1 \quad (13)$$

- $$R = \frac{n+1-2\sum_{i=1}^{s} M_i}{n-1} \qquad \text{(II)},$$

donde,

$$M_i = n_i \left[Q_i - \frac{(n_i - 1)\, x_i}{2A_n} \right],$$

- $$R^* = \frac{\sum_{i=1}^{s-1}(P_i - Q_i)}{\sum_{i=1}^{s-1} P_i} \quad \text{(I)}$$

Tanto la fórmula (13), como la (II) son fórmulas exactas. La fórmula (I) es una aproximación que aparece en muchos libros de textos, pero que en ocasiones tiene un comportamiento bastante deficiente.

3. Para datos agregados en intervalos:

Aproximando la verdadera curva de concentración, que es la poligonal dada para datos desagregados en frecuencias unitarias (o no unitarias, pues coinciden en ambos casos) por la poligonal obtenida al agregar los datos en intervalos.

- $$R' = \frac{\sum_{k=1}^{r}\left[S_k\,(N_{k-1} + N_k - 1)\right]}{(n-1)A_n} - 1 = \frac{ACA}{ACM} \quad (15)$$

- $IG_G = \frac{ACA}{1/2} = 2ACM = \frac{\sum_{k=1}^{r}\left[S_k\left(N_{k-1}+N_k\right)\right]}{nA_n} - 1$ (III)

- $R'' = \frac{1}{(n-1)A_n}\left[\sum_{k=1}^{r}\left[S_k\left(N_{k-1}+N_k-1\right)+\frac{1}{6}\left(n_k^{\,2}-1\right)\left(l_k-l_{k-1}\right)\right]\right]-1$ (17)

- $R^* = \frac{\sum_{i=1}^{r-1}\left(P_i-Q_i\right)}{\sum_{i=1}^{r-1}P_i}$.

Igual que en el caso de agrupación de frecuencias, la aproximación R^* que aparece en algunos libros de texto puede comportarse bastante mal.

La fórmula R" es una muy buena aproximación al valor verdadero de R, y es la que permanece más estable independientemente del número de clases que se consideren. La aproximación R' de R será peor a medida que la verdadera curva de concentración poligonal se aleje de la nueva forma poligonal que le hemos asignado; y por consiguiente, mientras más acentuados sean los cambios de pendiente de la curva de concentración verdadera, mientras menor sea el número de clases que hemos establecido y mientras mayor dispersión haya en el interior de cada clase.

Bibliografía.

1. Gini, Corrado. (1910), *Indici di Concentrazione e di dipendenza*, Biblioteca dell'Economista. Turín:Utet.

2. Gini, Corrado. (1912), "Variabilità e Mutabilità", *Studi Economico-Giuridici dell'Univ. Di Cagliari*, **3**, part 2, pp.1-158.

3. Gini, Corrado. (1914), "Sulla misura della concentrazione e della variabilità dei caractteri", *Atti del R. Istituto Veneto di Scienze, Lettere ed Arti*, Tomo **LXXIII**, pp. 1203-1248.