ACERCA DE "*SULLA MISURA DELLA CONCENTRAZIONE E DELLA VARIABILITÁ*" DE CORRADO GINI.

Basulto Santos, Jesús

Departamento de Economía Aplicada I.

Universidad de Sevilla.

e-mail: basulto@us.es

Romero García, José Enrique.

Departamento de Economía Aplicada I.

Universidad de Sevilla.

e-mail: romerogje@us.es

Resumen

Hemos realizado un análisis de este artículo de Corrado Gini (1884-1965) de 1914, principalmente de las secciones que contienen los resultados más interesantes. En el artículo, Gini 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. En la sección 6 del trabajo, Gini estudia la relación de su fórmula con la curva de concentración de Lorenz (1905), señalando la aproximación entre su fórmula y el doble del área de Lorenz a medida que el número de observaciones, n, crece; y en la sección 8 relaciona el doble del área de Lorenz con su formula, observando que se diferencian en el factor $\frac{n-1}{n}$. En la sección 9, Gini prueba la relación entre la medida *media de las diferencias sin reposición* con su *coeficiente de concentración.*

Palabras clave: Lorenz, dispersión, concentración, Gini.

1. Introducción.

Nuestro interés en Gini está justificado por ser un ejemplo de investigador en el que, en un primer período, trabajó la estadística como un instrumento y una metodología, para pasar, a partir de la primera guerra mundial, a trabajar la Estadística como Ciencia Estadística. Como señalamos, en la pequeña biografía que recogemos más abajo, los trabajos de Gini fueron realizados en los campos de la llamada ciencia social y de la estadística. También, Gini es de interés por contribuir a las primeras estimaciones de ciertas magnitudes económicas de los países o grupos de países, mucho antes de crearse la Contabilidad Nacional. Igualmente, Gini es reconocido como el inventor, en parte, del método estadístico para medir las paridades del poder de compra que usan los Institutos de Estadística Europeos. Gini también contribuyó a las largas discusiones sobre el uso del muestreo de poblaciones finitas, que tuvieron lugar en las distintas reuniones del Instituto Internacional de Estadística. Por último, la época que le tocó vivir en Italia, su preocupación por los problemas de la población y su concepción corporativista de la economía, plantean muchos interrogantes acerca de las relaciones de la Ciencia Estadística con la formación de los estados y sus sociedades. A continuación, como ya indicamos, recogemos una pequeña biografía de Gini.

Corrado Gini nació el 23 de mayo de 1884 en Motta di Livenza, cerca de Treviso, en el seno de una familia de terratenientes, y falleció en las primeras horas del 13 de marzo de 1965. Estudió Derecho, Estadística y Economía en la Universidad de Bologna, también estudió Matemáticas y Biología. Sus trabajos científicos fueron realizados, principalmente, en los campos de las ciencias sociales y la estadística. Gini estudió a Bernoulli, Lexis y Czuber, así como a Bodio, Messedaglia y Benini, maestros italianos de la ciencia estadística. En 1910 accedió a la cátedra de estadística en la Universidad de Cagliari, y el período hasta el final de la primera guerra mundial será el más importante por la relevancia de sus contribuciones a la metodología estadística.

En 1911 fue elegido miembro del Consejo Superior de Estadística, en 1913 Gini tomó posesión de la cátedra de estadística en Papua y en 1919 fue profesor en las Universidades de Cagliari y Padua, dando clases de Economía Política, Derecho

Constitucional, Demografía y Estadística Económica. En 1920 fundó METRON, la revista internacional de estadística, que dirigió hasta su muerte.

En 1923 se trasladó a la Universidad de Roma, comenzando un período de nueve años dedicado a actividades públicas; así puso en marcha la Escuela de Estadística para la formación de funcionarios en estadística, y en este mismo año de 1923 fundó la Facultad de Estadística, Demografía y Ciencias Actuariales. En 1926 fundó la revista *La Vita Economica Italiania* que fue cerrada en 1943, en 1929 Gini fundó el Comité Italiano para el estudio de los problemas de la población, y fundó la revista GENUS en 1934.

Gini fue uno de los más activos y distinguidos miembros del Instituto Internacional de Estadística, llegando a ser miembro honorario en 1939.

Gini, el mismo año de su muerte, recogió en su artículo "On the Characteristics of Italian Statistics" (1965) lo que para él era la Ciencia Estadística. Este artículo venía a completar a otro, "The Contributions of Italy to Modern Statistical Methods", de 1926. Destaquemos aquí algunas de sus frases, sin buscar mayor profundidad por no ser ahora nuestro objetivo.

"Una característica especial de la Estadística Italiana es describir cuantitativamente, por medio de series de constantes, los aspectos del fenómeno, colectivo de forma completa…"; "para mí, el principio fundamental…, en formular métodos estadísticos, es, primero, tener en cuenta la naturaleza del fenómeno al que son aplicados con relación al objeto de la investigación y sólo, en segundo lugar, considerar sus propiedades formales…"; "…muchos de los índices introducidos por la escuela italiana, no dependen del tipo de modelos que sigan los datos…".

Buscando aproximarnos al pensamiento de Gini, nos ha parecido que nuestro avance sería más seguro si nos centramos en uno de los índices más conocidos de Gini, en concreto en el índice o razón de concentración de Gini. Con el riesgo de perder la visión de conjunto, hemos estudiado el artículo "Sulla misura della concentrazione e della variabilità dei caractteri" de 1913-1914. Este artículo surgió como consecuencia del artículo "Indici di Concentrazione e di dipendenza" (1910) y la memoria "Variabilità e Mutabilità" de 1912.

A partir de aquí, recogemos en la sección 2 la razón de concentración para datos desagregados en frecuencias unitarias, generalizando dicha razón a datos agregados en frecuencias no unitarias, sección 3, y a datos agregados en intervalos, sección 4. En la sección 5, relacionamos la razón de concentración con la curva de Lorenz. Por último, en la sección 6, relacionamos la razón de concentración con la diferencias de media. No hemos incluido la discusión de medir la concentración con datos truncados.

Terminamos esta introducción indicando que hemos usado dos tipos de numeración de las fórmulas, una la ordinal habitual, (1), (2),...., que se corresponde con la dada por Gini, y otra, con números romanos, (I), (II),...., que es la usada por nosotros.

2. La Razón de Concentración, R.

En la sección 1 de su artículo de 1914, 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.

Define las variables $A_i = \sum_{k=1}^{i} a_k$, cantidad acumulada del carácter hasta el lugar i-ésimo. Con lo cual, la variable $Q_i = \frac{A_i}{A_n} = \frac{\sum_{k=1}^{i} a_k}{\sum_{k=1}^{n} a_k}$ constituye la proporción que representa la cantidad acumulada del carácter hasta el lugar i, A_i, respecto a la cantidad total presente del carácter. Mientras que la variable $P_i = \frac{i}{n}$ denota la proporción de observaciones hasta el lugar i respecto al número de observaciones totales.

Tomando dos rangos i,l, tal que $i<l$, se tiene que a_i a_l, y evidentemente, por las propiedades de la media aritmética,

$$\frac{1}{i}\sum_{k=1}^{i} a_k \le \frac{1}{l}\sum_{k=1}^{l} a_k ,$$

con lo que

$$\frac{\sum_{k=1}^{i} a_k}{\sum_{k=1}^{l} a_k} \le \frac{i}{l},$$

y, tomando $l=n$, se verifica que

$$\frac{\sum_{k=1}^{i} a_k}{\sum_{k=1}^{n} a_k} \le \frac{i}{n};$$

es decir,

$$Q_i \le P_i$$

Gini afirma que *"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 concentración del carácter será perfecta si sus intensidades son todas nulas, salvo una: $a_i=0$, $i=1,..,n-1$ y $a_n=A_n$. Dicho de otra forma, cuando $Q_i=0$, $i=1,..,n-1$; o también, cuando $P_i - Q_i = P_i$, $i=1,..,n-1$. La concentración será nula cuando $Q_i = P_i$, $i=1,..,n-1$; en tal caso se dirá que hay equidistribución del carácter.

A continuación, Gini se plantea buscar una medida o índice de concentración del carácter, a partir de las *n-1* desigualdades (4).

Es en este momento, cuando Gini hace referencia a su trabajo "*Indici di concentrazione e di dipendenza*", de 1910, y a varios trabajos donde se aplica el contenido de este último citado, como su "*Variabilitá e Mutabilitá*", de 1912.

En el trabajo de 1910, Gini define la medida de concentración basándose en la suposición de que la distribución del carácter sigue un modelo paramétrico determinado; en concreto supone la existencia de ciertas relaciones entre los pares $\{(P_i, Q_i), i=1,..,n\}$, como por ejemplo

$$1-P_i=(1-Q_i)^{\text{ä}}.$$

Por el contrario, en este trabajo de 1914, que es del que nos ocupamos en este artículo, Gini afirma: "*La presente nota tiene como objetivo el proponer una medida de concentración que sea independiente de la curva de distribución del carácter*".

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

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

Pero, ¿Cuál es el verdadero significado de R_i?.

$R_i = \frac{P_i - Q_i}{P_i}$ representa, en términos de proporción respecto a la intensidad media de todo el colectivo, la intensidad constante en que hay que aumentar la intensidad del carácter de cada uno de los P_i casos con intensidad más baja, para eliminar la desigualdad presente en los mismos respecto al conjunto del colectivo. En efecto:

Si en lugar de considerar los valores $a_1,..,a_i$, que constituyen la proporción P_i de observaciones, se consideran los valores transformados $b_1,..,b_i$, donde $b_j = a_j + R_i \cdot M \quad , j = 1,..,i$, y donde M es la media de todas las intensidades originales del carácter, se verifica que la media de esos nuevos valores transformados coincide con M, pues

$$M^b = \frac{\sum_{j=1}^{i} b_j}{i} = \frac{\sum_{j=1}^{i} \left(a_j + R_i \cdot M\right)}{i} =$$

$$= \frac{\sum_{j=1}^{i} a_j}{i} + \frac{MR_i \cdot i}{i} = \frac{\sum_{j=1}^{i} a_j}{i} + MR_i \, ,$$

de donde

$$M^b = \frac{\sum_{j=1}^{i} a_j}{i} + M \, \frac{P_i - Q_i}{P_i} = \frac{\sum_{j=1}^{i} a_j}{i} + M - M \, \frac{Q_i}{P_i} =$$

$$= \frac{\sum_{j=1}^{i} a_j}{i} + M - \frac{\sum_{j=1}^{N} a_j}{N} \, \frac{\sum_{j=1}^{i} a_j \Big/ \sum_{j=1}^{N} a_j}{i \big/ N} =$$

$$= \frac{\sum_{j=1}^{i} a_j}{i} + M - \frac{\sum_{j=1}^{i} a_j}{i} = M .$$

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 presentan 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 Gini señala que "*Como medida de la concentración 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,

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

Esa medida es la que Gini denomina *Razón de Concentración*.

También, 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.

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

La razón de concentración es por lo tanto, un índice de concentración que verifica las condiciones siguientes:

a) Crece al crecer la concentración,

b) Toma como valor más alto uno en el caso de máxima concentración

c) Toma su valor más pequeño, 0, en el caso de mínima concentración.

A continuación, veremos cómo se calcula en la práctica el valor de *R*. La formula que presentamos, equivalente a la mostrada en (11), es la que propone Gini al comienzo de su sección 3.

Dado que

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

se tiene que

$$R = \frac{\sum_{i=1}^{n-1}(P_i - Q_i)}{\sum_{i=1}^{n-1} P_i} = \frac{\sum_{i=1}^{n-1} P_i - \sum_{i=1}^{n-1} Q_i}{\sum_{i=1}^{n-1} P_i} = 1 - \frac{\sum_{i=1}^{n-1} Q_i}{\sum_{i=1}^{n-1} P_i} = 1 - \frac{\sum_{i=1}^{n-1} A_i / A_n}{(n-1)/2} = 1 - \frac{2}{(n-1)A_n}\sum_{i=1}^{n-1} A_i \ ,$$

y obtenemos la siguiente formulación para *R*:

$$R = 1 - \frac{2}{(n-1)A_n}\sum_{i=1}^{n-1} A_i \ . \quad (12)$$

Si ahora tenemos en cuenta que

$$\sum_{i=1}^{n-1} A_i = \sum_{i=1}^{n-1} \sum_{k=1}^{i} a_k \ ,$$

y que esa expresión , observando la tabla siguiente,

i	$A_i = \sum_{k=1}^{i} a_k$	
1	$A_1 = \sum_{k=1}^{1} a_k$	a_1

2	$A_2=\sum_{k=1}^{2}a_k$	a_1+a_2
3	$A_3=\sum_{k=1}^{3}a_k$	$a_1+a_2+a_3$
$\vdots$	$\vdots$	$\vdots$
n-1	$A_{n-1}=\sum_{k=1}^{n-1}a_k$	$a_1+a_2+a_3+..+a_{n-1}$

coincide con

$$\sum_{i=1}^{n-1}A_i=\sum_{i=1}^{n-1}\sum_{k=1}^{i}a_k=(n-1)\,a_1+(n-2)\,a_2+..+(n-(n-1))a_{n-1}=\sum_{i=1}^{n-1}(n-i)\,a_i\ ,$$

se cumplirá que

$$R=1-\frac{2}{(n-1)\,A_n}\sum_{i=1}^{n-1}A_i=\frac{(n-1)\,A_n-2\sum_{i=1}^{n-1}A_i}{(n-1)\,A_n}=$$

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

con lo que

$$R=\frac{n\sum_{i=1}^{n}a_i-\sum_{i=1}^{n}a_i-2\sum_{i=1}^{n-1}na_i+2\sum_{i=1}^{n-1}ia_i}{(n-1)\,A_n}=$$

$$=\frac{n\sum_{i=1}^{n}a_i-2\sum_{i=1}^{n}a_i+\sum_{i=1}^{n}a_i-2\sum_{i=1}^{n-1}na_i+2\sum_{i=1}^{n-1}ia_i}{(n-1)\,A_n};$$

y de ahí,

$$R=\frac{n\sum_{i=1}^{n}a_i+\sum_{i=1}^{n}a_i-2\sum_{i=1}^{n-1}na_i+2\left(\sum_{i=1}^{n-1}ia_i-\sum_{i=1}^{n}a_i\right)}{(n-1)A_n}=$$

$$=\frac{n\sum_{i=1}^{n}a_i+\sum_{i=1}^{n}a_i-2\sum_{i=1}^{n-1}na_i+2\left(\sum_{i=1}^{n}ia_i-na_n-\sum_{i=1}^{n}a_i\right)}{(n-1)A_n}=$$

$$=\frac{n\sum_{i=1}^{n}a_i+\sum_{i=1}^{n}a_i-2n\sum_{i=1}^{n-1}a_i-2na_n+2\left(\sum_{i=1}^{n}ia_i-\sum_{i=1}^{n}a_i\right)}{(n-1)A_n}=$$

$$=\frac{n\sum_{i=1}^{n}a_i+\sum_{i=1}^{n}a_i-2n\sum_{i=1}^{n}a_i+2\left(\sum_{i=1}^{n}ia_i-\sum_{i=1}^{n}a_i\right)}{(n-1)A_n}.$$

Por tanto,

$$R=\frac{(n-2n)\sum_{i=1}^{n}a_i+\sum_{i=1}^{n}a_i+2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n}=$$

$$=\frac{(n-2n+1)\sum_{i=1}^{n}a_i+2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n}=$$

$$=\frac{-(n-1)A_n+2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n},$$

de donde,

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

3. Agregación en frecuencias.

Gini, continuando con su sección 3, se plantea obtener una expresión de su 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 (él 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.

Observando la siguiente tabla de correspondencias:

$a_1,..,a_{n_1}$	$a_{n_1+1},..,a_{n_1+n_2}$	$a_{n_2+1},..,a_{n_2+n_3}$	...	$a_{n_{s-1}+1},..,a_{n_{s-1}+n_s}$
$a_1,..,a_{N_1}$	$a_{N_1+1},..,a_{N_2}$	$a_{N_2+1},..,a_{N_3}$	...	$a_{N_{s-1}+1},..,a_{N_s}$
x_1	x_2	x_3	...	ls

se tiene que

$$\sum_{i=1}^{n}(i-1)a_i = \sum_{i=1}^{N_1}(i-1)a_i + \sum_{i=N_1+1}^{N_2}(i-1)a_i + \cdots + \sum_{i=N_{s-1}+1}^{N_s}(i-1)a_i \qquad \text{(I)}$$

se podrá poner en este caso como:

$$\begin{aligned}\sum_{i=1}^{n}(i-1)a_i &= \sum_{i=1}^{N_1}(i-1)a_i + \sum_{i=N_1+1}^{N_2}(i-1)a_i + \cdots + \sum_{i=N_{s-1}+1}^{N_s}(i-1)a_i = \\ &= x_1\sum_{i=1}^{N_1}(i-1) + x_2\sum_{i=N_1+1}^{N_2}(i-1) + \cdots + x_s\sum_{i=N_{s-1}+1}^{N_s}(i-1) = \\ &= \sum_{l=1}^{s} x_l \sum_{i=N_{l-1}+1}^{N_l}(i-1).\end{aligned}$$

Luego,

$$\sum_{i=1}^{n}(i-1)a_i = \sum_{l=1}^{s} x_l \sum_{i=N_{l-1}+1}^{N_l}(i-1);$$

y, en consecuencia, la expresión (12 bis), $R = \frac{2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n} - 1$, se reduce a:

$$R=\frac{2\sum_{l=1}^{s}x_l\sum_{i=N_{l-1}+1}^{N_l}(i-1)}{(n-1)A_n}-1.$$

Pero como,

$$\sum_{i=N_{l-1}+1}^{N_l}(i-1)=(N_{l-1}+1-1)+(N_{l-1}+2-1)+\cdots+(N_{l-1}+n_l-1)=$$

$$=\frac{(N_{l-1}+1-1)+(N_{l-1}+n_l-1)}{2}n_l=\frac{(N_{l-1}+N_l-1)}{2}n_l;$$

es decir,

$$\sum_{i=N_{l-1}+1}^{N_l}(i-1)=\frac{(N_{l-1}+N_l-1)}{2}n_l, \quad \text{(II)}$$

se tiene que:

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

Para el caso de frecuencias unitarias, n_l=1, esta última fórmula coincide con la expresión (12 bis), ya que se tiene que

$s=n$, $x_l=a_l$ y, a su vez,: $N_{l-1}+N_l-1=(N_l-n_l)+N_l-1=(N_l-1)+N_l-1=2(N_l-1)$.

Para casos de frecuencias no unitarias, es claro que la fórmula (13) no coincide con la (12 bis).

En muchos manuales de estadística, se utiliza la fórmula (11) de Gini con datos agregados en frecuencias; y es claro que (11) es igual a (12 bis) y que esta no coincide con (13) en estos casos de datos agregados en frecuencias no unitarias.

Una explicación de este error puede deberse, de alguna manera, al hecho que en el libro de Gini "*Curso de Estadística*", de 1935, traducido por el estadístico y economista J.A. Vandellos, sólo se recogía la formula (11).

4. Agregación en intervalos.

Continuando con la sección 3 del artículo de Gini, Gini afronta la extensión de 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 expresión $2\sum_{i=1}^{n}(i-1)a_i$ que aparece en la fórmula (12 bis), define 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 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.

Para cada uno de los intervalo I_k, $k=1,..,r$ se considera:

- Para cada uno de los n_k valores a_i, $i=N_{k-1}+1,..N_{k-1}+n_k$, que caen dentro del intervalo I_k, las diferencias entre ellos y su valor medio en dicho intervalo, $\frac{S_k}{n_k}$:

$$\boldsymbol{d}_{ki} = a_i - \frac{S_k}{n_k}, \; i=N_{k-1}+1,..N_{k-1}+n_k,$$

con lo que

$$a_i = \boldsymbol{d}_{ki} + \frac{S_k}{n_k} \quad . \qquad \text{(III)}$$

- Igualmente, para cada uno de los n_k rangos i, $i=N_{k-1}+1,..N_{k-1}+n_k$, que caen dentro del intervalo I_k, se consideran las diferencias entre los valores i-1 y el valor medio de estas expresiones en dicho intervalo,

$$\frac{\sum_{i=N_{k-1}+1}^{N_k}(i-1)}{n_k} = \frac{\frac{(N_{k-1}+N_k-1)}{2}n_k}{n_k} = \frac{(N_{k-1}+N_k-1)}{2};$$

es decir, las diferencias

$$e_{ki} = (i-1) - \frac{(N_{k-1} + N_k - 1)}{2},$$

con lo que

$$(i-1) = e_{ki} + \frac{(N_{k-1} + N_k - 1)}{2}. \qquad \text{(IV)}$$

Como la suma de las desviaciones de los valores de una magnitud respecto a su media es cero, se tendrá que para cada uno de los intervalo I_k, k=1,..,r, se verifica:

$$\sum_{i=N_{k-1}+1}^{N_k} \boldsymbol{d}_{ki} = 0$$

$$\sum_{i=N_{k-1}+1}^{N_k} \boldsymbol{e}_{ki} = 0\,.$$

Por otro lado,

$$\sum_{i=1}^{n} (i-1)\, a_i = \sum_{i=1}^{N_1} (i-1)\, a_i + \sum_{i=N_1+1}^{N_2} (i-1)\, a_i + \cdots + \sum_{i=N_{r-1}+1}^{N_r} (i-1)\, a_i = \sum_{k=1}^{r} \sum_{i=N_{r-1}+1}^{N_r} (i-1)\, a_i\ ,$$

luego:

$$2\sum_{i=1}^{n} (i-1)\, a_i = \sum_{k=1}^{r} 2 \sum_{i=N_{r-1}+1}^{N_r} (i-1)\, a_i\ , \qquad \text{(V)}$$

y, si sustituimos en la expresión el valor obtenido en (III):

$$2\sum_{i=N_{k-1}+1}^{N_k} (i-1)\, a_i = 2\sum_{i=N_{k-1}+1}^{N_k} (i-1)\left(d_{ki} + \frac{S_k}{n_k} \right) = 2\sum_{i=N_{k-1}+1}^{N_k} (i-1)\, d_{ki} + 2\frac{S_k}{n_k} \sum_{i=N_{k-1}+1}^{N_k} (i-1)\,. \qquad \text{(VI)}$$

Pero, por (II):

$$\sum_{i=N_{k-1}+1}^{N_k} (i-1) = \frac{(N_{k-1}+N_k-1)}{2} n_k \,,$$

y, por otro lado, por (IV):

$$\begin{aligned}
\sum_{i=N_{k-1}+1}^{N_k} (i-1) d_{ki} &= \sum_{i=N_{k-1}+1}^{N_k} \left(e_{ki} + \frac{(N_{l-1}+N_l-1)}{2} \right) d_{ki} = \\
&= \sum_{i=N_{k-1}+1}^{N_k} e_{ki} d_{ki} + \frac{(N_{l-1}+N_l-1)}{2} \sum_{i=N_{k-1}+1}^{N_k} d_{ki} = \\
&= \sum_{i=N_{k-1}+1}^{N_k} e_{ki} d_{ki} \,,
\end{aligned}$$

luego,

$$\sum_{i=N_{k-1}+1}^{N_k} (i-1) d_{ki} = \sum_{i=N_{k-1}+1}^{N_k} e_{ki} d_{ki} \,.$$

Y, sustituyendo en (VI), se tiene:

$$2 \sum_{i=N_{k-1}+1}^{N_k} (i-1) a_i = 2 \sum_{i=N_{k-1}+1}^{N_k} e_{ki} d_{ki} + S_k \left(N_{k-1} + N_k - 1 \right).$$

Por consiguiente, la expresión (V), queda:

$$\begin{aligned}
2\sum_{i=1}^{n} (i-1) a_i &= \sum_{k=1}^{r} 2 \sum_{i=N_{r-1}+1}^{N_r} (i-1) a_i = \sum_{k=1}^{r} \left[2 \sum_{i=N_{k-1}+1}^{N_k} e_{ki} d_{ki} + S_k \left(N_{k-1} + N_k - 1 \right) \right] = \\
&= \sum_{k=1}^{r} \left[S_k \left(N_{k-1} + N_k - 1 \right) \right] + 2 \sum_{k=1}^{r} \left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki} d_{ki} \right];
\end{aligned}$$

es decir,

$$2\sum_{i=1}^{n} (i-1) a_i = \sum_{k=1}^{r} \left[S_k \left(N_{k-1} + N_k - 1 \right) \right] + 2 \sum_{k=1}^{r} \left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki} d_{ki} \right].$$

Con esto, la fórmula (12 bis) se expresa en la forma:

$$R=\frac{2\sum_{i=1}^{n}(i-1)a_i}{(n-1)A_n}-1=\frac{\sum_{k=1}^{r}\left[S_k\left(N_{k-1}+N_k-1\right)\right]+2\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k}e_{ki}d_{ki}\right]}{(n-1)A_n}-1.$$

En conclusión, para el caso de valores agrupados en intervalos:

$$R=\frac{\sum_{k=1}^{r}\left[S_k\left(N_{k-1}+N_k-1\right)\right]+2\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k}e_{ki}d_{ki}\right]}{(n-1)A_n}-1. \qquad (14)$$

Como el término $\frac{2\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k}e_{ki}d_{ki}\right]}{(n-1)A_n}$ es pequeño, la formula anterior conduce a esta otra fórmula aproximada:

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

Y, si además n es grande, esa formula (15) puede aproximarse por

$$R'_2=\frac{\sum_{k=1}^{r}\left[S_k\left(N_{k-1}+N_k\right)\right]}{nA_n}-1, \qquad (15\text{ bis})$$

verificándose que

$$R'_2=\frac{n-1}{n}R'.$$

Como tanto los d_{ki}, como los e_{ki} crecen dentro de cada clase I_k, se tiene que $\sum_{i=N_{k-1}+1}^{N_k}e_{ki}d_{ki}>0$, y por tanto $\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k}e_{ki}d_{ki}\right]>0$; de donde se obtiene que:

R’<R. (16)

A continuación Gini busca aproximar el término

$$\frac{2\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki}d_{ki}\right]}{(n-1)\,A_n}, \qquad \text{(VII)}$$

y para ello, la primera hipótesis que considera es 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.

Si llamamos b, aunque en realidad habría que llamarle b_k, a la diferencia entre los términos de esa progresión aritmética, podemos considerar varias posibilidades:

1. Que ninguno de los dos extremos de cada intervalo sea un valor observado, con lo que se tendrá que $b = \frac{l_k - l_{k-1}}{n_k + 1}$
2. Que los dos extremos de cada intervalo sea un valor observado, con lo que se tendrá que $b = \frac{l_k - l_{k-1}}{n_k - 1}$
3. Que solo uno de los dos extremos de cada intervalo sea un valor observado, por ejemplo el superior, con lo que se tendrá que $b = \frac{l_k - l_{k-1}}{n_k}$

Por simplificar los cálculos, nos ceñiremos a este caso (que es compatible con la forma $I_k=(l_{k-1}, l_k]$ para los intervalos) , con lo cual el primer valor que aparece en el intervalo I_k será $l_{k-1}+b$, y el último será $l_{k-1}+ n_k\, b$, que coincidirá con l_k.

Calcularemos los valores tanto de d_{ki}, como de los e_{ki}, para posteriormente sustituirlos en la expresión

$$\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki}d_{ki}\right], \qquad \text{(VIII)}$$

que forma parte de (VII).

Por (III), $\boldsymbol{d}_{ki} = a_i - \frac{S_k}{n_k}$. Pero, a su vez, de la tabla siguiente,

$I_k=(l_{k-1}, l_k]$, $i=N_{k-1}+1$
$a_i=l_{k-1}+b$ $a_{i+1}=l_{k-1}+2b$ $\vdots$ $a_{i+(n_k-1)}=l_{k-1}+n_k\, b$

se deduce que

$$S_k= l_{k-1}n_k+(1+2+..+n_k)b= l_{k-1}n_k + b\, n_k(n_k+1)/2,$$

y, por tanto,

$$\frac{S_k}{n_k} = l_{k-1} + \frac{b(n_k+1)}{2},$$

luego

$$\boldsymbol{d}_{ki} = a_i - \frac{S_k}{n_k} = a_i - l_{k-1} - \frac{b(n_k+1)}{2},\ i=N_{k-1}+1,..N_{k-1}+n_k,$$

y, en consecuencia,

$$\begin{aligned}\boldsymbol{d}_{k,N_{k-1}+l} &= a_{N_{k-1}+l} - l_{k-1} - \frac{b(n_k+1)}{2} = \\ &= (l_{k-1}+lb) - l_{k-1} - \frac{b(n_k+1)}{2} = b\left(l - \frac{n_k+1}{2}\right),\ l = 1,..,n_k.\end{aligned}$$

Es decir,

$$d_{k,N_{k-1}+l} = b\left(l - \frac{n_k+1}{2}\right), l = 1,..,n_k . \qquad \text{(IX)}$$

Análogamente, por (IV):

$$e_{ki} = (i-1) - \frac{(N_{k-1}+N_k-1)}{2} = (i-1) - \frac{(N_{k-1}+N_{k-1}+n_k-1)}{2} =$$
$$= (i-1) - N_{k-1} - \frac{n_k-1}{2},$$

con lo que

$$e_{ki} = (i-1) - N_{k-1} - \frac{n_k-1}{2}, \; i=N_{k-1}+1,..N_{k-1}+n_k ;$$

y, por tanto,

$$e_{k,N_{k-1}+l} = (N_{k-1}+l-1) - N_{k-1} - \frac{n_k-1}{2} = l - \frac{n_k+1}{2}, l = 1,..,n_k .$$

Es decir,

$$e_{k,N_{k-1}+l} = l - \frac{n_k+1}{2}, l = 1,.., n_k . \qquad \text{(X)}$$

Por (IX) y (X), se verifica que:

$$e_{k,N_{k-1}+l} = be_{k,N_{k-1}+l}, l = 1,..,n_k .$$

Luego,

$$\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki}d_{ki}\right] = \sum_{k=1}^{r}\left[\sum_{l=1}^{n_k} e_{k,N_{k-1}+l}d_{k,N_{k-1}+l}\right] =$$
$$= \sum_{k=1}^{r}\left[\sum_{l=1}^{n_k} e_{k,N_{k-1}+l} \cdot be_{k,N_{k-1}+l}\right] = \sum_{k=1}^{r} b\left[\sum_{l=1}^{n_k} e^2{}_{k,N_{k-1}+l}\right];$$

es decir,

$$\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki}d_{ki}\right] = \sum_{k=1}^{r} b\left[\sum_{l=1}^{n_k} e^2{}_{k,N_{k-1}+l}\right] . \qquad \text{(XI)}$$

Pero, como por (X):

$$e_{k,N_{k-1}+l} = l - \frac{n_k+1}{2} = (l-1) + 1 - \frac{n_k+1}{2} =$$
$$= (l-1) - \frac{n_k-1}{2}, l = 1,..,n_k,$$

se tiene que

$$e_{k,N_{k-1}+l} = (l-1) - \frac{n_k-1}{2}, l = 1,..,n_k;$$

luego,

$$\sum_{l=1}^{n_k} e^2{}_{k,N_{k-1}+l} = \sum_{l=1}^{n_k}\left((l-1) - \frac{n_k-1}{2}\right)^2 = \sum_{l=1}^{n_k}\left((l-1)^2 + \left(\frac{n_k-1}{2}\right)^2 - 2(l-1)\frac{n_k-1}{2}\right) =$$
$$= \sum_{l=1}^{n_k}(l-1)^2 + \sum_{l=1}^{n_k}\left(\frac{n_k-1}{2}\right)^2 - (n_k-1)\sum_{l=1}^{n_k}(l-1).$$

Y, como sabemos que $\sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6}$, se tendrá que:

$$\sum_{l=1}^{n_k} e^2{}_{k,N_{k-1}+l} = \sum_{l=1}^{n_k}(l-1)^2 + \sum_{l=1}^{n_k}\left(\frac{n_k-1}{2}\right)^2 - (n_k-1)\sum_{l=1}^{n_k}(l-1) =$$
$$= \frac{(n_k-1)n_k\left[\,2(n_k-1)+1\right]}{6} + \left(\frac{n_k-1}{2}\right)^2 n_k - (n_k-1)\frac{(n_k-1)\,n_k}{2} =$$
$$= \frac{(n_k-1)\,n_k\left[4n_k - 2 + 3n_k - 3 - 6n_k + 6\right]}{12} = \frac{(n_k-1)\,n_k\left[n_k+1\right]}{12} =$$
$$= \frac{n_k\left(n_k{}^2-1\right)}{12};$$

es decir,

$$\sum_{l=1}^{n_k} e^2{}_{k,N_{k-1}+l} = \frac{n_k\left(n_k{}^2-1\right)}{12}.$$

Y sustituyendo esa expresión en (XI), se tiene que:

$$\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki}d_{ki}\right]=\sum_{k=1}^{r} b\left[\sum_{l=1}^{n_k} e^2{}_{k,N_{k-1}+l}\right]=\sum_{k=1}^{r}\frac{n_k\left(n_k{}^2-1\right)}{12}b\,. \qquad \text{(XII)}$$

Por tanto,

$$\frac{2\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki}d_{ki}\right]}{(n-1)A_n}=\frac{2\sum_{k=1}^{r}\frac{n_k\left(n_k{}^2-1\right)}{12}b}{(n-1)A_n}=\frac{1}{6(n-1)A_n}\sum_{k=1}^{r} n_k\left(n_k{}^2-1\right)b\,.$$

Así pues, si tomamos b bajo el supuesto que dijimos, $b=\frac{l_k-l_{k-1}}{n_k}$,se tendrá que

$$\frac{2\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki}d_{ki}\right]}{(n-1)A_n}=\frac{1}{6(n-1)A_n}\sum_{k=1}^{r} n_k\left(n_k{}^2-1\right)\frac{l_k-l_{k-1}}{n_k}=$$
$$=\frac{1}{6(n-1)A_n}\sum_{k=1}^{r}\left(n_k{}^2-1\right)\left(l_k-l_{k-1}\right).$$

Es decir,

$$\frac{2\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki}d_{ki}\right]}{(n-1)A_n}=\frac{1}{6(n-1)A_n}\sum_{k=1}^{r}\left(n_k{}^2-1\right)\left(l_k-l_{k-1}\right), \qquad \text{(XIII)}$$

y sustituyendo esa expresión en (14), se tiene que, en el caso de agrupación en intervalos, y bajo el supuesto establecido de progresión aritmética, la expresión aproximada de R, es:

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

Es decir,

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

Igualmente, si n es suficientemente elevado, esa formula puede aproximarse por:

$$R_2'' = \frac{1}{nA_n}\left[\sum_{k=1}^{r}\left[S_k\left(N_{k-1}+N_k\right)+\frac{1}{6}n_k^2\left(l_k-l_{k-1}\right)\right]\right]-1, \qquad \text{(17 bis)}$$

existiendo la siguiente relación entre ambas:

$$R_2'' = \frac{n-1}{n}R'' + \frac{l_r - l_0}{6(n-1)A_n}\ .$$

Obsérvese, que las fórmulas (15), (15 bis), (17) y (17 bis) suponen que conocemos los totales, S_k, las frecuencias absolutas, n_k, y los extremos de cada una de las clases I_k. Pero,

a) En algunas aplicaciones, no se conoce el extremo inferior de la primera clase y tampoco el extremo superior de la última clase. En estos casos, una hipótesis que podemos usar es:

$$\frac{S_k}{n_k} = \frac{l_{k-1}+l_k}{2}\ .$$

b) En otras ocasiones, sucede que, para cada clase, solamente conocemos sus frecuencias absolutas y desconocemos los valores de S_k, en tal caso podemos usar también la hipótesis dada en a), con lo que

$$S_k = \frac{l_{k-1}+l_k}{2}n_k$$

$$A_n = \sum_{k=1}^{r} S_k = \frac{1}{2}\sum_{k=1}^{r}\left(l_{k-1}+l_k\right)n_k\ .$$

Para tales situaciones, Gini, sustituyendo en (17) las expresiones anteriores, haya su última aproximación a R:

$$R''' = \frac{\sum_{k=1}^{r}\left[\left(N_{k-1}+N_k-1\right)n_k\left(l_k+l_{k-1}\right)+\frac{1}{3}\left(n_k^2-1\right)\left(l_k-l_{k-1}\right)\right]}{(n-1)\sum_{k=1}^{r}\left(l_{k-1}+l_k\right)n_k}-1, \qquad (18)$$

que a su vez, cuando n es muy elevado, puede aproximarse por:

$$R_2'''' = \frac{\sum_{k=1}^{r}\left[(N_{k-1}+N_k)n_k\,(l_k+l_{k-1})+\frac{1}{3}\left(n_k^{\;2}\right)(l_k-l_{k-1})\right]}{n\sum_{k=1}^{r}(l_{k-1}+l_k)n_k}-1\,. \quad \text{(18 bis)}$$

A partir de este momento, Gini dedica su sección 4 a recoger varias aplicaciones de R en diferentes ciencias.

En su sección 5, Gini plantea la situación en donde el carácter cuantitativo tiene un número v de nulos y un número n de valores positivos, estableciendo que

$$R_t = R_p \cdot m + (1-m), \quad m = \frac{n-1}{n+v-1}\,, \qquad (19)$$

donde R_p es la razón de concentración calculada en los valores exclusivamente positivos del carácter, y donde R_t es la razón de concentración para el conjunto de todos los valores del carácter.

5. Relación entre *R* y la Curva de Concentración de Lorenz.

En la sección 6 del artículo, Gini, trabajando con datos desagregados con frecuencias absolutas unitarias, prueba que $\frac{\sum_{i=1}^{n-1}(P_i-Q_i)}{n}$ se aproxima al área de concentración, área entre la recta de equidistribución y la curva de concentración de Lorenz, cuando n es suficientemente elevado.

Igualmente, demuestra que $\frac{\sum_{i=1}^{n-1}P_i}{n}$ se aproxima, para n suficientemente grande, al área del triangulo formado por la recta de equidistribución y la base, y el lado derecho del cuadrado que delimita la curva de concentración.

Así pues, Gini prueba que, para n suficientemente grande, la razón de concentración, $R = \frac{\sum_{i=1}^{n-1} (P_i - Q_i)}{\sum_{i=1}^{n-1} P_i}$, se aproxima al cociente entre las dos áreas citadas anteriormente.

En sus secciones 7 y 8, Gini propone usar métodos aproximados para calcular el área de concentración. Pero recalca, que para al cálculo aritmético la fórmula de R'' que aparece en (17) es una buena aproximación al valor de R.

En la sección 8, construye una curva de concentración con datos agregados en intervalos de clase. A continuación une los puntos de cambio de la curva de concentración por segmentos; así aproxima la curva de concentración por una poligonal y luego calcula la aproximación al área de concentración como el área existente entre la recta de equidistribución y la poligonal.

Veamos seguidamente la fórmula que 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:

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

Ahora, para calcular una media aproximada de la concentración Gini divide 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 es:

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

expresión que se relaciona con la fórmula de *R*', hallada en (15), que aproximaba a la razón de concentración *R*, pues coincide con el valor $\frac{n-1}{n}R'$.

Obviamente, 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):

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

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

En consecuencia, la aproximación *R*' de *R*, que realizamos en (15) 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. Por tanto, 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.

6. Relación entre *R* y la Diferencia de Medias.

En su sección 9, Gini prueba, partiendo de la fórmula (12), para datos desagregados en frecuencias absolutas unitarias, que la *diferencia media relativa sin reemplazamiento* coincide con la razón de concentración. Es decir,

$$R = \frac{\Delta}{2M},$$

donde M es la media aritmética, y donde,

$$\Delta = \frac{\sum^{n}\sum^{n}\left|a_i - a_j\right|}{n(n-1)} = \frac{2\sum_{i=1}^{\frac{n+1}{2}}(n+1-2i)(a_{n+1-i} - a_i)}{n(n-1)},$$

y donde el valor (n+1)/2 hay que cambiarlo por n/2 en caso que n sea par.

En su sección 10, Gini relaciona, para el caso de datos agregados en intervalos de clase, la diferencia de medias con la aproximación de R que notamos por R', y que aparecía en la formula (15).

Si los n valores del carácter cuantitativo se agrupan en r clases, denotamos por Δ_k la diferencia media para los n_k valores del carácter dentro de la clase I_k, y por Δ' la diferencia media de la agrupación que se obtiene en el colectivo en conjunto al sustituir los n_k valores de cada una de las clase I_k por su correspondiente media aritmética, se verifica que:

$$\Delta' = \Delta - \frac{\sum_{k=1}^{r} n_k (n_k - 1)\Delta_k}{n\ (n-1)}. \qquad (23)$$

Ahora, Gini indica, aunque no lo demuestra, que de forma análoga a como demostró que $R = \frac{\Delta}{2M}$, se prueba que

$$R' = \frac{\Delta'}{2M}.$$

Veámoslo por otro camino.

Para datos agregados en frecuentas absolutas, es conocido que también se verifica la relación $R = \frac{\Delta}{2M}$; luego, al sustituir los n_k valores de cada una de las clase I_k por su correspondiente media aritmética, denominémosla x_k, se verifica que la razón de concentración de estos nuevos datos, llamémosla R^*, vendrá dada por

$$R^* = \frac{\Delta'}{2M}. \qquad \text{(XIV)}$$

Pero, si agregamos en frecuencias absolutas, por (13), el valor de R^* será

$$R^* = \frac{\sum_{k=1}^{r} \left(N_{k-1} + N_k - 1 \right) x_k n_k}{(n-1) A_n} - 1 \ ,$$

que coincide con $R' = \frac{\sum_{k=1}^{r} \left[S_k \left(N_{k-1} + N_k - 1 \right) \right]}{(n-1) A_n} - 1$, la expresión aproximada de R al caso de agrupación de intervalos, hallada en (15), pues $x_k n_k = S_k$, al ser x_k la media aritmética de la clase I_k, luego, $R' = R^*$, y ,por (XIV), se verifica:

$$R' = R^* = \frac{\Delta'}{2M};$$

es decir,

$$R' = \frac{\Delta'}{2M},$$

que es lo que queríamos probar.

Si ahora volvemos a (23), $\Delta' = \Delta - \frac{\sum_{k=1}^{r} n_k (n_k - 1) \Delta_k}{n\ (n-1)}$, y dividimos los dos miembros de la igualdad entre 2M, se tiene que:

$$\frac{\Delta'}{2M} = \frac{\Delta - \frac{\sum_{k=1}^{r} n_k (n_k - 1) \Delta_k}{n\ (n-1)}}{2M} \Rightarrow R' = \frac{\Delta}{2M} - \frac{\sum_{k=1}^{r} n_k (n_k - 1) \Delta_k}{2Mn\ (n-1)} = R - \frac{\sum_{k=1}^{r} n_k (n_k - 1) \Delta_k}{2Mn\ (n-1)} \ ,$$

luego,

$$R' = R - \frac{\sum_{k=1}^{r} n_k (n_k - 1) \Delta_k}{2M \cdot n \ (n-1)}, \qquad (24)$$

$$R = R' + \frac{\sum_{k=1}^{r} n_k (n_k - 1) \Delta_k}{2M \cdot n \ (n-1)}. \qquad (XV)$$

Por (14) y (15), se tenía que:

$$R = R' + \frac{2\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki} d_{ki}\right]}{(n-1)\, A_n},$$

por tanto, teniendo en cuenta (XV), se verifica que:

$$\frac{2\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki} d_{ki}\right]}{(n-1)\, A_n} = \frac{\sum_{k=1}^{r} n_k (n_k - 1) \Delta_k}{2Mn \ (n-1)}. \qquad (XVI)$$

Calculemos seguidamente una aproximación al término

$$\frac{\sum_{k=1}^{r} n_k (n_k - 1) \Delta_k}{2M \cdot n \ (n-1)}.$$

Si ahora, suponemos que los valores del carácter cuantitativo, dentro de cada clase, siguen una progresión aritmética de diferencia $b_k = \frac{l_k - l_{k-1}}{n_k}$, Gini, obtiene que

$$\Delta_k = \frac{n_k + 1}{3} b_k,$$

y, por tanto,

$$\frac{\sum_{k=1}^{r} n_k\,(n_k-1)\Delta_k}{2M\cdot n\,(n-1)}=\frac{\sum_{k=1}^{r} n_k\,(n_k-1)\frac{n_k+1}{3}b_k}{2\frac{A_n}{n}\cdot n\,(n-1)}=\frac{\sum_{k=1}^{r} n_k\,(n_k-1)(n_k+1)\frac{l_k-l_{k-1}}{n_k}}{6A_n(n-1)}=$$

$$=\frac{\sum_{k=1}^{r}\left(n_k{}^2-1\right)(l_k-l_{k-1})}{6A_n(n-1)};$$

es decir,

$$\frac{\sum_{k=1}^{r} n_k\,(n_k-1)\Delta_k}{2M\cdot n\,(n-1)}=\frac{\sum_{k=1}^{r}\left(n_k{}^2-1\right)(l_k-l_{k-1})}{6A_n(n-1)}, \qquad \text{(XVII)}$$

y, sustituyendo en (XV), se obtiene que, en el caso de agrupación en intervalos, y bajo el supuesto de progresión aritmética, la expresión aproximada para R es:

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

que coincide con la dada en (17).

Además, en esta situación, se verifica, por (XVI) y (XVII), que

$$\frac{2\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki}d_{ki}\right]}{(n-1)\,A_n}=\frac{\sum_{k=1}^{r} n_k\,(n_k-1)\,\Delta_k}{2Mn\,(n-1)}=\frac{\sum_{k=1}^{r}\left(n_k{}^2-1\right)(l_k-l_{k-1})}{6A_n\,(n-1)};$$

y por tanto,

$$\frac{2\sum_{k=1}^{r}\left[\sum_{i=N_{k-1}+1}^{N_k} e_{ki}d_{ki}\right]}{(n-1)\,A_n}=\frac{\sum_{k=1}^{r}\left(n_k{}^2-1\right)(l_k-l_{k-1})}{6A_n\,(n-1)},$$

valor que también habíamos obtenido anteriormente en (XIII)

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.

4. Gini, Corrado. (1926), "The Contributions of Italy to Modern Statistical Methods", *J.R. Statist. Soc.*, **89**, pp. 703-724.

5. Gini, Corrado. (1965), "On the Characteristics of Italian Statistics", *J.R. Statist. Soc. Serie A,* Volume **128**, 89-109.

6. Gini, Corrado, (1935), *Curso de Estadística.* Con un apéndice Matemático por Luigi Galvani. Traducido del italiano por J.A. Vandellós. Editorial Labor.