CONSIDERACIONES SOBRE EL ÍNDICE DE GINI PARA MEDIR LA CONCENTRACIÓN DEL INGRESO

DIVISIÓN DE ESTADÍSTICA Y PROYECCIONES ECONÓMICAS
CEPAL

ÍNDICE

I. INTRODUCCIÓN

El grado de desigualdad económica existente en una sociedad y su evolución en el tiempo son temas que mantienen el interés permanente de la opinión pública y de los especialistas en el estudio del bienestar colectivo. Por su parte, en la literatura que aborda el análisis de la distribución del excedente de la economía, se han propuesto diferentes medidas que pretenden sintetizar esta variable, con el objeto de efectuar comparaciones intertemporales y entre países, a la vez de permitir asignar un valor absoluto a la desigualdad y derivar conclusiones sobre el nivel de concentración del ingreso en una población determinada.

En este contexto, es práctica común que los académicos y especialistas en el tema, así como los funcionarios de oficinas de gobierno y de organismos internacionales, y el público en general, se preocupen por conocer la manera en que los distintos grupos de hogares participan en la formación del ingreso nacional.

Diversos son los análisis que se hacen sobre el tema, así como los procedimientos metodológicos que se aplican para evaluar el grado de inequidad que existe en una sociedad. A lo largo de la historia del análisis económico, se han propuesto diversos indicadores para el estudio de la desigualdad; sin embargo, parece existir consenso en el hecho de que el indicador que ha tenido mayor aceptación en los trabajos empíricos es el denominado coeficiente de concentración de Gini. Este índice, de fácil interpretación, es una referencia común en los debates sobre el bienestar y la equidad; además, la opinión pública está muy pendiente de su evolución para sancionar el funcionamiento de los gobiernos en materia de desigualdad y sus efectos en el nivel de vida de la población.

El objetivo de esta nota es pasar revista a las nociones fundamentales que se asocian con la medición de la desigualdad económica, así como analizar los aspectos teóricos y conceptuales que sustentan la construcción de diversos indicadores que se utilizan en la investigación empírica.

En la segunda sección se aborda la medición de la desigualdad, estableciendo las diferencias que existen entre los índices de carácter estadístico y aquellos que se derivan de las funciones de bienestar social. El tercer apartado se aboca al estudio de la curva de Lorenz como medida de equidad, señalando su importancia en el análisis de la distribución del ingreso, así como su relevancia para la construcción de indicadores agregados como el índice de Gini que resume todas las características de la distribución en un solo valor. Posteriormente, la cuarta parte del trabajo se centra en la medición de la concentración del ingreso a partir del Indice de Gini, analizando por separado las fórmulas alternativas que se aplican cuando se dispone de datos agrupados e información desagregada. Asimismo, se analizan aspectos relevantes asociados al concepto de ingreso que se considera más apropiado para aproximar la desigualdad, el efecto del número de grupos en el valor del indicador, así como la variable relevante para el ordenamiento de la información y aquella que se utiliza para acumular el presupuesto de las familias. Finalmente, la quinta sección hace revisión de los procedimientos de descomposición sugeridos para el índice de Gini concluyendo que este indicador presenta problemas cuando se intenta conocer la variabilidad generada por las diferencias de ingresos al interior de los grupos y por la formación de los mismos, mientras que en el último apartado se pasa revista a las fuentes de información que habitualmente se emplean para el estudio de la distribución del ingreso.

II. LA MEDICIÓN DE LA DESIGUALDAD

Un índice de desigualdad es una medida que resume la manera como se distribuye una variable entre un conjunto de individuos. En el caso particular de la desigualdad económica, la medición se asocia al ingreso (o al gasto) de las familias o personas. Así, si $\mathbf{y_1, y_2, \ldots, y_n}$ representan los ingresos de un grupo de n individuos, el indicador de desigualdad se construye como función de las observaciones: $\mathbf{I(y_1, y_2, \ldots, y_n)}$.

Una primera clasificación de los indicadores de desigualdad que se encuentra en la literatura los agrupa como sigue: **medidas positivas**, que son aquellas que no hacen referencia explícita a ningún concepto de bienestar social, y medidas **normativas**, que sí están basadas en una función de bienestar. Al primer grupo pertenecen los índices estadísticos que tradicionalmente se utilizan para analizar la dispersión de una distribución de frecuencias, en tanto que hay diversas medidas normativas que se han propuesto para el estudio de la concentración del ingreso y la salud[1].

Dependiendo del indicador seleccionado, se define la norma o parámetro con la cual se compara la distribución del ingreso observada. En el caso de los estadísticos que comúnmente se utilizan para estudiar la dispersión de una variable **(varianza, desviación estándar** y **coeficiente de variación)**, el valor de referencia está representado por el promedio de la variable de análisis ($\bar{y}$). (Véase el Recuadro 1)

Pese a la sencillez de cálculo de las medidas de dispersión, no es muy común que éstas se utilicen para el estudio de la desigualdad, debido a que no satisfacen algunas propiedades teóricas que deben cumplir los buenos indicadores para el análisis de la distribución del ingreso (para el detalle de cuáles son algunas de estas propiedades, véase el Recuadro 4). A modo de ejemplo, debe señalarse que ninguna de ellas satisface la condición de independencia de escala, ya que el valor del indicador se altera cuando las observaciones se multiplican por una constante positiva:

$$\mathbf{I\,[cy_1, cy_2, \ldots, cy_n] \neq I\,[y_1, y_2, \ldots, y_n]\ \forall\ y_i}\ \text{ y }\ \mathbf{c > 0} \qquad \mathbf{(1)}$$

Por el contrario, una de las medidas probablemente más utilizadas es el denominado **Coeficiente de Gini (CG)**. Este indicador, que se clasifica entre las medidas estadísticas para el análisis de la distribución del ingreso, no utiliza como parámetro de referencia el ingreso medio de la distribución –a diferencia de la desviación media, la varianza y el coeficiente de variación–, dado que su construcción se deriva a partir de la curva de Lorenz.

Recuadro 1

Medidas de Desigualdad Positivas

Existe un conjunto de expresiones estadísticas tradicionalmente utilizadas para conocer la dispersión de una determinada variable en un conjunto de datos, las cuales se han incorporado en los trabajos que se abocan al estudio de la desigualdad económica.

Para el caso en que se analiza la distribución del ingreso, considere que $\mathbf{y_1}$, $\mathbf{y_2}$, …, $\mathbf{y_n}$ representan los ingresos de n individuos o familias. Además, que:

$\mu = \sum_{i=1}^{n} \frac{y_i}{n}$ representa el valor medio del ingreso y $n\mu = \sum_{i=1}^{n} y_i$, el ingreso total de la población.

La manera más elemental de estudiar la dispersión de la variable ingreso, $\mathbf{y_i}$, en una distribución es comparando los valores extremos observados. Así, se define el **rango**:

$$\mathbf{Rango = [\max(y_i) - \min(y_i)]} \qquad \mathbf{(1)}$$

1 Otra clasificación que resulta útil es la siguiente: estadísticos tradicionales, índice de Gini, medidas basadas en la entropía, y los índices basados en funciones de bienestar social.

Cuando se considera la distancia que existe entre el rango y la media de la distribución μ, se obtiene el denominado **rango relativo**:

$$\textbf{Rango Relativo} = \frac{[\max(y_i) - \min(y_i)]}{\mu} \qquad (2)$$

La principal debilidad de las medidas anteriores es que están basadas únicamente en las observaciones extremas y, por lo tanto, ignoran información relevante del resto de los datos analizados.

Otra manera de estudiar la dispersión de una variable en una distribución es mediante la comparación de la suma del valor absoluto de todas las diferencias respecto al valor medio, con relación al valor total de la variable. De esta forma, se define lo que se conoce como la **desviación media relativa**:

$$\textbf{Desviación Media Relativa} = \frac{\sum_{i=1}^{n} |\mu - y_i|}{n\mu} \qquad (3)$$

Para que se cumpla que el rango de variación de la ecuación (3) esté entre 0 y 1, la desviación media relativa se expresa como:

$$\textbf{DMR} = \frac{\sum_{i=1}^{n} |\mu - y_i|}{2n\mu} \qquad (3a)$$

El principal problema de esta medida es que no es sensible a las transferencias de ingresos que se puedan efectuar entre personas que están del mismo lado con respecto al ingreso medio de la distribución. Para resolver este problema, es posible considerar la sumatoria de las desviaciones con respecto a la media y elevarlas al cuadrado, de tal suerte que las diferencias se acentúen en la medida en que una observación $\mathbf{y_i}$ se aleja del valor medio de ingresos de la distribución μ. Conforme a lo anterior, se define la **varianza**:

$$\textbf{Varianza (V)} = \frac{\sum_{i=1}^{n} (\mu - y_i)^2}{n} \qquad (4)$$

La varianza cumple con el principio de transferencias de ingresos. En 1920, H. Dalton –continuando con el razonamiento introducido por Pigou (1912)– señaló que este principio es lo mínimo que debe cumplir cualquier medida de desigualdad, por lo que en la literatura se la denomina condición de Pigou-Dalton (Sen, 1996).

Entre las características de la varianza, podemos señalar que cualquier transferencia de ingresos de una persona pobre a una más rica necesariamente incrementará su valor, debido a que aumentará la distancia entre la observación que se ve favorecida y el valor medio de la distribución. Además, este indicador no es independiente de las unidades de medida, ya que cuando los ingresos se incrementan en una proporción $c > 0$, entonces la varianza se incrementa en c^2. Si bien la varianza cumple con la condición de Pigou-Dalton, el efecto de cualquier transferencia de una persona con ingreso y a otra con ingreso $(y+k)$ es el mismo, independientemente del valor de y.

Otra medida a la que se recurre con frecuencia en el análisis es la **desviación estándar**, la cual se define simplemente como la raíz cuadrada de la varianza.

$$\textbf{Desviación Estándar (DE)} = \sqrt{V} \qquad (5)$$

Al igual que la varianza, la desviación estándar depende del valor medio de la variable. Esto podría ocasionar que una distribución tenga una menor varianza que otra, a pesar de presentar una mayor variación relativa, si es que el ingreso medio de la primera distribución es menor que el de la segunda.

Este problema se resuelve utilizando el **coeficiente de variación**:

$$\textbf{Coeficiente de Variación (CV)} = \frac{\sqrt{V}}{\mu} \qquad \textbf{(6)}$$

El coeficiente de variación es una medida independiente del nivel medio de ingresos μ, y es, además, sensible a cualquier transferencia de ingresos en la distribución. Sin embargo, la sensibilidad ante transferencias no depende del valor de *y*. Es común que en el trabajo empírico se encuentren mediciones efectuadas con el $\mathbf{CV^2}$, debido a la relación que tiene con los índices de entropía.

Si se desea que una medida de desigualdad otorgue mayor importancia a las transferencias de ingresos que se generan en la parte baja de la distribución, se recomienda el uso de la transformación logarítmica. Esto da lugar a dos medidas comúnmente utilizadas, la **varianza de los logaritmos** y la **desviación estándar de los logaritmos**:

$$\textbf{Varianza de los Logaritmos (VL)} = \frac{\sum\left(\log\mu - \log y_i\right)^2}{n} \qquad \textbf{(7)}$$

La forma de esta función hace que las observaciones con ingresos bajos pesen más en el índice que los ingresos elevados, lo cual es deseable para algunos fines; además, este indicador se puede descomponer en forma aditiva. Sin embargo, la VL tiene un comportamiento no deseado en la parte superior de la distribución, lo que ocasiona que las transferencias de los muy ricos a los menos ricos aumenten la concentración en lugar de reducirla.

$$\textbf{Desviación Estándar de los Logaritmos (DEL)} = \sqrt{\frac{\sum\left(\log\mu - \log y_i\right)^2}{n}} \qquad \textbf{(8)}$$

Esta transformación, a diferencia de la varianza y la desviación estándar, tiene la ventaja que elimina los problemas asociados a las unidades de medida. Además, a las transferencias que se realizan en la parte baja de la distribución se les asigna mayor ponderación que a las que se efectúan en los niveles altos de ingreso.

III. LA CURVA DE LORENZ

En el estudio de la desigualdad, se dispone de diversos métodos para describir la forma en que se distribuye el ingreso entre los diferentes grupos de individuos en una sociedad: los **diagramas de dispersión**, los **indicadores de desigualdad** y los **ordenamientos de la información**.

Elaborar un diagrama para visualizar la distribución del ingreso resulta una opción extremadamente útil para el análisis de la desigualdad, ya que permite identificar ciertos aspectos de la forma de la distribución que de otra manera no sería posible apreciar. En la literatura se consignan al menos cuatro posibles alternativas para generar ordenamientos de datos, aunque sólo nos ocuparemos en el análisis de las dos que se emplean con mayor frecuencia. Estas son las distribuciones de frecuencias, la curva de Lorenz, los denominados diagramas de desfile propuestos por Pen y la transformación logarítmica.

La **distribución de frecuencias** es la manera más intuitiva de ordenar las observaciones, al agrupar el ingreso de los individuos en diferentes intervalos de clase y observar la concentración de observaciones que se forman al interior de cada uno de ellos. Sin embargo, con este tipo de representación gráfica no se muestran en forma adecuada las colas de la distribución. Adicionalmente, las observaciones que se agrupan al interior de los intervalos quedan representadas por el punto medio o marca de clase, lo cual necesariamente conduce a pérdida de información.

En la medida de que el número de grupos es pequeño, la pérdida de información es mayor y este tipo de gráficos no aporta nada al conocimiento de la forma en que se distribuyen las observaciones. Por el contrario, hacer uso de esta opción supone formar un número suficiente de intervalos de tal manera que se observen las características relevantes de la población asociadas a sus principales medidas de tendencia central y a la dispersión que existe en torno a ellas.

Tal vez la forma más habitual de representar la desigualdad sea a partir de la **Curva de Lorenz**. Esta medida fue propuesta en 1905 con el propósito de ilustrar la desigualdad en la distribución de la salud y, desde su aparición, su uso se ha popularizado entre los estudiosos de la desigualdad económica.

En términos simples, la curva de Lorenz representa el porcentaje acumulado de ingreso (**$\%Y_i$**) recibido por un determinado grupo de población (**$\%P_i$**) ordenado en forma ascendente de acuerdo a la cuantía de su ingreso (**$y_1 \leq y_2 \leq ..., \leq y_n$**).

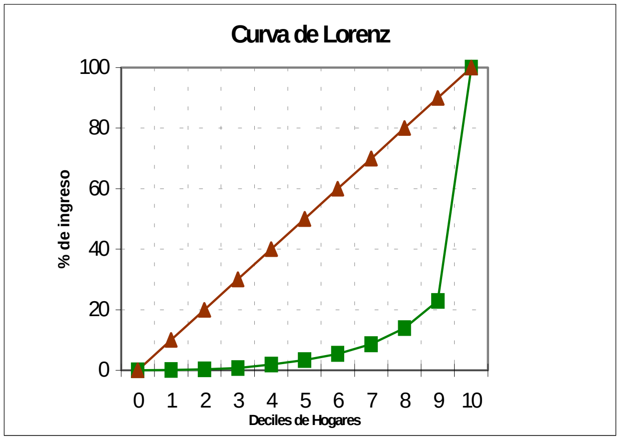
La construcción de esta curva se efectúa como se muestra a continuación: Supóngase que se tienen n individuos ordenados en forma creciente respecto al valor de sus ingresos $y_1 \leq y_2, \ldots, \leq y_{n.,}$ y se forman g grupos de igual tamaño, llamados percentiles (quintiles o deciles son las formas más comunes de ordenar a la población). La curva de Lorenz se define como la relación que existe entre las proporciones acumuladas de población (%**P_i**) y las proporciones acumuladas de ingreso (%**Y_i**).

De esta manera, en caso que a cada porcentaje de la población le corresponda el mismo porcentaje de ingresos ($P_i=Y_i$; $\forall$ i), se forma una línea de 45^0. Como puede observarse en la figura 1, esta línea divide en dos partes iguales el cuadrado de lado uno que se forma al graficar las proporciones acumuladas de personas en el eje horizontal (P_i) y de ingresos en el vertical (Y_i). Dicha diagonal corresponde a lo que Lorenz definió como la **línea de equidad perfecta** y denota, por ende, ausencia de desigualdad.

En la gráfica 1 se observan tres elementos: la línea de equidistribución, la curva correspondiente a la distribución empírica formada a partir de las parejas (P_i,Y_i), o curva de Lorenz, y el área entre las dos líneas, denominada **área de concentración**. A partir de la relación entre esta última y la línea de igualdad perfecta, es posible derivar diversos indicadores que se utilizan para evaluar la concentración del ingreso.

Debe observarse que la curva de Lorenz siempre se ubica por debajo de la diagonal en la medida que los ingresos de los individuos se hayan ordenado en forma creciente, y por encima en el caso opuesto. En la medida que la curva de Lorenz se aproxime a la diagonal, se estaría observando una situación de mayor igualdad, mientras que cuando se aleja, la desigualdad se incrementa. Obviamente, el punto (0,0) significa que el 0% de la población tiene el 0% del ingreso, en tanto que en el extremo opuesto el 100% de la población concentra todo el ingreso.

FIGURA 1



Desde el punto de vista empírico, la curva de Lorenz se construye graficando los porcentajes acumulados de ingresos que reciben los distintos grupos de la población (Y_i), con la única condición de que estos se definan con la misma amplitud, a efectos de evitar problemas asociados al número de observaciones agrupadas en cada intervalo. En el ámbito del análisis de la distribución del ingreso de las familias, es común que la curva de Lorenz se construya a partir de datos agrupados en subconjuntos iguales de tamaño 10%, denominados deciles de hogares, y que se utilicen diversos conceptos de ingreso para efectuar el ordenamiento de las observaciones.

De esta manera, la inclinación de cada segmento de la curva se determina a partir del cociente que se forma al dividir el porcentaje de ingreso apropiado por un determinado segmento de la población, por el ingreso medio de la distribución. Así, en la medida que la inclinación que tenga el segmento sea más pronunciada, mayor será también la proporción de ingreso que retiene ese grupo. Es evidente que cuando la inclinación de un determinado segmento coincide con la de la diagonal, el ingreso de ese grupo es igual al promedio de la distribución.

El hecho de que para la construcción de la curva intervengan únicamente los porcentajes de población e ingresos, aísla el efecto del ingreso total y, por lo tanto, ésta sólo refleja la estructura y forma de la distribución. A su vez, cuando dos curvas de Lorenz se intersectan, es posible demostrar que una distribución puede obtenerse de la otra a partir de transferencias regresivas o progresivas de ingresos. Consecuentemente, en esta situación no se puede concluir cuál de las distribuciones comparadas tiene un mayor grado de desigualdad[2].

Recuadro 2

Los Indices de Desigualdad Basados en Funciones de Bienestar Social

El primer trabajo que proponía el uso de funciones de bienestar social para medir la desigualdad se atribuye a Dalton (1920). En esa investigación, el autor propuso medir la proporción del bienestar que se pierde debido a la presencia de una inequitativa distribución del ingreso entre las personas. Utilizando una función de utilidad aditiva, separable, simétrica y estrictamente cóncava del ingreso, $u(y_i)$, definió lo que en la literatura se conoce como el **Indice de Dalton**:

2 Esta inconsistencia se resuelve a partir de la construcción de las denominadas curvas de Lorenz generalizadas. Dichas curvas se construyen multiplicando la función de Lorenz por el ingreso medio de la distribución (μ), de manera que el indicador ya no varíe entre cero y uno, sino que el límite superior esté determinado por el valor del ingreso medio (μ).

Sean $\mathbf{y_1, y_2, \ldots, y_n}$ los ingresos observados en una muestra de tamaño n, y sea μ el promedio de ingresos de la distribución; entonces el Indice de Dalton se expresa por medio de:

$$D = 1 - \sum_{i=1}^{n} \frac{u(y_i)}{nu(\mu)} \qquad \textbf{(1)}$$

Debido a que la función propuesta es cóncava, el índice siempre asume valores positivos, salvo cuando todas las observaciones tienen el mismo nivel de ingreso y D toma el valor de 0. La utilidad práctica de este indicador ha sido ampliamente discutida en la literatura. Para cierto tipo de funciones, como las logarítmicas y las hiperbólicas, el Indice de Dalton cumple con la propiedad de que los incrementos iguales no dependen del valor medio de la distribución; sin embargo, se observa que el valor del indicador es invariante a transformaciones lineales positivas de la función de utilidad, lo cual le ha valido la crítica de los especialistas.

Basándose en estos argumentos, y redefiniendo el pensamiento de Dalton, en 1970 Atkinson propuso una familia de índices normativos que resultan invariantes a cambios de escala y a transformaciones lineales positivas de la función de utilidad. La sugerencia de este autor se basa en el criterio de definir para cada población el nivel de ingreso equivalente $\mathbf{y_e}$, de tal forma que si cada individuo recibiera ese monto de recursos, el bienestar total sería el mismo para toda la población. Es decir, que $\mathbf{W(y_e\, e_n) = W(y_1, y_2, \ldots, y_n)}$, en donde $\mathbf{e_n}$ representa un vector unitario de dimensión n.

El índice de Atkinson se calcula de la manera siguiente:

$$\mathbf{A = 1 - y_e / \mu} \qquad \textbf{(2)}$$

Cuando se trabaja con una función de utilidad estrictamente cóncava, se cumple que $\mathbf{y_e < \mu}$; por lo tanto, A será siempre positivo y asumirá el valor 0 únicamente cuando todos los individuos tengan el mismo nivel de ingreso.

Una de las interpretaciones más comunes a partir de los valores que asume el índice de Atkinson es la siguiente: Suponga que el valor del índice para una población determinada es de 0.35. Esta situación implica que si los recursos se distribuyeran de manera igualitaria entre todos los individuos, con tan sólo el 65% del total de ingresos se podría garantizar el nivel de equidad observado, lo cual significa que el 35% restante se despilfarra a causa de la inequidad que existe en la distribución del ingreso.

La propuesta operativa para el cálculo del índice de Atkinson supone la existencia de funciones de bienestar aditivamente separables, a partir de las cuales propone la siguiente familia de índices, consistentes con el ordenamiento que genera la curva de Lorenz:

$$\mathbf{A_\alpha (y) = 1 - [\, \textstyle\sum^{n}_{i=1} (y_i/\mu)^{1-\alpha}\,]^{1/(1-\alpha)} \;;\; \alpha > 0 \text{ y } \alpha \neq 1} \qquad \textbf{(3)}$$

$$\mathbf{A_\alpha (y) = 1 - \textstyle\prod^{n}_{i=1} (y_i/\mu)^{1/n} \;;\; \alpha = 1} \qquad \textbf{(4)}$$

Para garantizar que esta familia de índices satisfaga las propiedades deseables, se les debe imponer a las funciones de utilidad ciertas restricciones con las que muchos investigadores no están de acuerdo. Sin embargo, el índice de Atkinson es ampliamente utilizado en los estudios de la desigualdad, debido a que permite captar en forma adecuada lo que sucede en la parte baja de la distribución. En la medida que se incrementa el valor de α –parámetro asociado con la aversión social a la desigualdad–, las transferencias entre los más pobres se ponderan en mayor proporción. En el caso en que $\alpha \to \infty$, la función de bienestar coincide con la propuesta por Ralphs( ), en cuyo caso sólo se estarían analizando las transferencias que recibe el individuo más pobre de toda la distribución.

La utilidad práctica de considerar distintos valores del parámetro para la familia de índices de Atkinson, se advierte al observar que es muy útil para el análisis generar una serie de indicadores con el fin de observar qué sucede con el nivel de la desigualdad, en la medida que en la expresión se le otorga mayor importancia a las familias ubicadas en la parte baja de la distribución del ingreso.

Por otra parte, si una curva queda totalmente contenida en la otra (salvo en los valores extremos) puede afirmarse, sin ambigüedad, que aquella que se ubica más cerca de la diagonal presenta una distribución más igualitaria, en cuyo caso se dice que domina en el orden de Lorenz. En términos algebraicos lo anterior se expresa diciendo que la distribución "x" domina a "y" en el sentido de Lorenz, si se cumple que:

$$\sum_{i=1}^{n} y_i^x \geq \sum_{i=1}^{n} y_i^y \qquad \textbf{(2)}$$

donde y_i representa el ingreso de la familia

Las curvas de Lorenz permiten comparar niveles de desigualdad relativa, cumpliendo el principio de preferencia por la igualdad y simetría, sin tener que calcular ningún índice adicional. Finalmente, cabe señalar que un índice de desigualdad será consistente con el orden de Lorenz, cuando su valor sea menor en la distribución dominante que en la dominada.

IV. EL COEFICIENTE DE DESIGUALDAD DE GINI

a) Datos Desagregados

Sin lugar a dudas, uno de los índices más utilizados en el estudio de la desigualdad es el Coeficiente de Concentración de Gini (CG). Existen diversas formas de derivar la expresión algebraica que se usa para su cálculo, y también es posible deducirlo desarrollando un procedimiento geométrico a partir de la curva de Lorenz.

Gini (1912) definió su conocida medida de desigualdad en los términos siguientes:

$$CG = \frac{1}{2\mu}\left[\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i - y_j\right|}{n(n-1)}\right] = \frac{1}{2\mu}\Delta \qquad \textbf{(3)}$$

en donde Δ representa la media aritmética de las n(n-1) diferencias absolutas de las observaciones y 2μ es el valor máximo que asume Δ cuando un individuo concentra todo el ingreso.

Posteriormente, en 1914 Gini propuso un nuevo indicador de concentración que se define como 1 menos dos veces el área de la curva de Lorenz, y demostró que era equivalente a la que había presentado en un trabajo previo y que corresponde a la definida en la expresión (3).

$$\mathbf{CG = 1 - 2\,F(y)} \qquad \textbf{(4)}$$

donde F(y) representa la curva de Lorenz; es decir, la proporción de individuos o familias que tienen ingresos acumulados menores o iguales a y.[3]

La derivación de (4) se basa en el razonamiento de que si el área de concentración de la figura 1 se divide por el área del triángulo que se ubica debajo de la línea de igualdad perfecta, se obtiene una

---

3 En términos de la teoría estadística sobre la distribución del ingreso, F(*Y*) representa la probabilidad de que la variable aleatoria *Y* asuma un valor menor o igual a *y*, [F(*Y*) = P(*Y* ≤ *y*)].

medida de concentración. Debido a que el área bajo la línea de equidistribución es igual a ½ y el área total es igual a la unidad, entonces a partir del razonamiento anterior se deriva la expresión (5).

Considerando la gráfica 1, el coeficiente de Gini se define como el cociente de las diferencias entre la línea de equidistribución y los valores de la curva de Lorenz. Existen varias maneras de derivar algebraicamente el índice de Gini, y una de ellas demuestra que es exactamente igual a la mitad de la **diferencia media relativa** (**DMR**), la que se define como la media aritmética de las diferencias absolutas entre todos los pares de ingresos.

$$CG = \frac{DMR}{2} = \frac{\dfrac{\sum_{i,j}\left(y_i - y_j\right)}{\bar{y}} \dfrac{1}{n^2}}{2} = \frac{\sum_{i,j}\left(y_i - y_j\right)}{2n^2\bar{y}} \qquad \textbf{(5)}$$

Recuadro 3

Indices de Desigualdad Basados en la Entropía

Utilizando los conceptos de la teoría de la información, en la literatura económica se han propuesto algunos indicadores para medir la desigualdad, basados en medidas de entropía. Tal vez el que se conoce en forma más amplia es el denominado índice de Theil.

Suponga que una variable aleatoria puede asumir los valores $\mathbf{y_1, y_2, \ldots, y_n}$, con probabilidades $\mathbf{p_1, p_2, \ldots, p_n}$ mayores o iguales a cero y cuya suma es igual a la unidad. Si se efectúa una selección aleatoria, mientras menor sea la probabilidad de selección de la observación $\mathbf{y_i}$, mayor será la relevancia de la selección efectuada. La idea anterior se utiliza para construir indicadores que le asignen más importancia a las familias o individuos que menor proporción del ingreso retengan. En este caso, se define $\mathbf{p_i = y_i / (n\mu)}$, en donde $\mathbf{p_i}$ se interpreta como la proporción de ingreso retenida por el i-ésimo individuo.

Con base en lo anterior, la entropía de orden α de Rényi (1965) genera una medida de la cantidad de información contenida en una variable aleatoria con densidad de probabilidad discreta $\mathbf{p = [p_1, p_2, \ldots , p_n]}$:

$$\mathbf{I_\alpha (p) = [1 / (1-\alpha)] \log [\textstyle\sum^n_{i=1} p_i\, \alpha] \ ; \alpha > 0 \text{ y } \alpha \neq 1} \qquad \textbf{(1)}$$

Theil (1967) definió como medida de desigualdad del ingreso, la diferencia entre la entropía que se deriva de la situación de igualdad perfecta y la calculada para la distribución empírica, lo cual se interpreta como la entropía que se genera debido a que el ingreso no se distribuye en forma igualitaria.

La expresión práctica que se emplea para calcular el índice de Theil se presenta a continuación:

$$\mathbf{T_\alpha (y) = 1/n \textstyle\sum^n_{i=1} \{ y_i / [\mu \log (y_i/\mu)] \}; \alpha = 1} \qquad \textbf{(2)}$$

$$\mathbf{T_\alpha (y) = 1/n \textstyle\sum^n_{i=1} [\log (\mu/y_i)] \ ; \alpha = 0} \qquad \textbf{(3)}$$

En el caso de (2), el índice de Theil le otorga mayor importancia a los ingresos que se ubican en la parte baja de la distribución, mientras que en (1) se le asigna aún más peso a los hogares de menores ingresos.

Además de satisfacer las propiedades de independencia de la media y de invarianza respecto a replicaciones de población, el índice de Theil cumple con la propiedad de descomposición aditiva, característica poco frecuente entre los indicadores comúnmente utilizados para medir la desigualdad. Esta propiedad lo hace especialmente atractivo, ya que siempre resulta importante conocer qué porcentaje de la desigualdad está explicado por la inequidad que se genera entre los grupos formados y cuál proviene de las diferencias de ingresos al interior de los mismos.

A diferencia de las medidas estadísticas que tradicionalmente son utilizadas para estudiar la dispersión de un conjunto de datos –donde la media es el parámetro de localización que se utiliza como referencia–, el índice de Gini se construye al comparar la distribución empírica que se forma con los datos observados y la línea de igualdad perfecta que supone la distribución teórica derivada de la curva de Lorenz.

Nuevamente, supóngase que se tiene información sobre n individuos ordenados en forma creciente respecto al valor de sus ingresos: $\mathbf{y_1 \leq y_2, \ldots, \leq y_n}$. Además, se construye la distribución de frecuencias relativas simple y acumulada de la población bajo estudio (p_i , P_i), así como de la variable a distribuir (en este caso el ingreso del hogar y_i e Y_i). Nótese que, debido al ordenamiento de los datos, se cumple que $Y_i \leq Y_{i+1}$.

Conforme a lo anterior, el coeficiente de desigualdad de Gini se basa en la suma de las diferencias (P_i - Y_i). Con el propósito de estandarizar su recorrido al intervalo [0,1], la expresión anterior se divide entre $\sum^{n-1}_{i=1} P_i$, dando paso a una de las fórmulas que comúnmente se utilizan para calcular el **índice de Gini cuando se dispone de datos no agrupados:**

$$CG = \frac{\sum_{i=1}^{n-1}\left(P_i - Y_i\right)}{\sum_{i=1}^{n-1} P_i} \qquad \textbf{(6)}$$

El valor mínimo del indicador es cero, y se obtiene cuando $P_i=Y_i \ \forall_i$, mientras que asume el valor 1 cuando $Y_1=Y_2= \ldots=Y_{n-1}=0$, ya que en este caso el individuo n concentra todo el ingreso Y_n.

Otras expresiones que se utilizan frecuentemente para el cálculo del coeficiente de desigualdad de Gini son:

$$CG = \frac{1}{2n^2}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|y_i - y_j\right|}{\bar{y}} \qquad \textbf{(7)}$$

$$CG = 1 + \frac{1}{n} - \frac{2}{n^2}\,y\left(y_1 + 2y_2 + \ldots + ny_n\right) \qquad \textbf{(7a)}$$

El CG induce un ordenamiento de las observaciones que es consistente con la curva de Lorenz, y además asigna mayor peso a las observaciones que se encuentran en la parte baja de la distribución, independientemente de cuáles sean sus valores. No obstante, una de las principales desventajas de este indicador es que las transferencias unitarias de ingreso entre individuos reciben más peso en la medida que haya más observaciones en la zona donde se enfectúan. Por lo tanto, la ponderación máxima corresponde a transferencias próximas a la moda y lejos de los extremos de la distribución.

b) Datos Agrupados

En el trabajo empírico es habitual que el cálculo de los índices de desigualdad se efectúe a partir de las observaciones agrupadas, debido a que no resulta práctico comparar conjuntos de datos que pueden llegar a tener tamaños muy distintos. En ese sentido, lo común es que se decida agrupar las observaciones en subconjuntos de igual tamaño de modo que se faciliten tanto los cálculos como las comparaciones entre grupos.

Sin embargo, esta manera de proceder conduce necesariamente a la pérdida de información, debido a que los valores individuales de las observaciones serán sustituidos por algún valor que represente al intervalo en que se encuentran agrupados (por ejemplo, la marca de clase).

De acuerdo a lo anterior, el procedimiento que se utiliza para el cálculo del índice de Gini para datos agrupados es el siguiente:

i) Ordenar los hogares en forma ascendente conforme a su ingreso.

ii) Definir intervalos de igual tamaño (por ejemplo, deciles de hogares).

iii) Construir la distribución de frecuencias relativas, simple y acumulada, de la variable a distribuir (ingreso), así como de la población que se desea estudiar. Si se opta por la formación de deciles de hogares, cada grupo deberá concentrar el 10% de las observaciones, tal como se muestra en el cuadro 1.

Recuadro 4

Algunas Propiedades Deseables de los Indicadores para Medir la Desigualdad

Diversos investigadores han definido un conjunto de propiedades deseables que deben satisfacer los buenos indicadores utilizados para medir la desigualdad económica, algunas de las cuales se presentan a continuación:

1. La medida de desigualdad propuesta debe ser invariante a **transformaciones proporcionales o cambios de escala**. Si **I** representa el indicador de desigualdad calculado a partir de un vector de ingresos $\mathbf{y = (y_1, y_2, \ldots, y_n)}$, entonces el valor de $\mathbf{I(y) = I(y_1, y_2, \ldots, y_n)}$ no debe modificarse ante una nueva medición que utilice $\mathbf{y^* = c\ y}$, donde **c** es una constante mayor que 0. Lo anterior significa que la medida de desigualdad no debe variar si se efectúa una transformación de la escala en que se mide la variable de interés; por ejemplo, si se pasa de miles de pesos a millones en la medición de la concentración del ingreso personal, el indicador propuesto no debe alterarse por ese hecho.

En términos matemáticos esta propiedad se conoce como la propiedad de homogeneidad en grado cero. Los índices que satisfacen esta condición son particularmente útiles para realizar comparaciones intertemporales e internacionales de la desigualdad, ya que resultan independientes de las unidades monetarias en que se mida.

2. **Condición de Pigou-Dalton**. Esta condición sugiere que, si se genera una transferencia de ingresos de los hogares ubicados en la parte superior de la distribución hacia las familias o personas ubicadas en los primeros percentiles, el indicador propuesto debe reflejar necesariamente una caída en el nivel de concentración. En el estudio de la pobreza, esta propiedad equivale al axioma de transferencia definido por Sen (1976) para los buenos índices de pobreza.

Cowell (1995) ha denominado a esta propiedad como principio débil de transferencias, ya que no se especifica nada acerca de la magnitud en la caída de la desigualdad. En contraposición, este mismo autor define el principio fuerte de transferencias, que resulta esencialmente atractiva cuando se analizan aspectos vinculados a la descomposición de los índices.

Se dice que una medida satisface este principio, si el monto en la reducción de la inequidad a consecuencia de una transferencia de ingresos depende sólo de la distancia entre los ingresos, sin importar cuál sea la posición de los individuos en la distribución. Al imponer esta condición se asume que todos los individuos tienen necesidades comparables ignorando las diferencias que existen en su edad, composición familiar, etc. Entonces, un indicador será una buena medida de desigualdad si se observa una caída sistemática en su valor cuando nos aproximamos a la equidistribución. En caso contrario, su valor se debería incrementar, toda vez que las transferencias se generen de la parte más baja de la distribución hacia los grupos más favorecidos. Según Kakwani (1980), toda medida de desigualdad que se construya como la media aritmética de una función estrictamente convexa en el ingreso satisface esta propiedad.

Supóngase que el índice de concentración genera un valor x. Entonces, si se observan redistribuciones sucesivas, necesariamente habrá cambios en el nivel de desigualdad hasta alcanzar el valor $\mathbf{l_m}$ como límite superior, en el caso de que todo el ingreso lo concentre una persona. En caso contrario, nos deberemos aproximar a $\mathbf{l_i=0}$ en la medida que se realicen transferencias hacia la parte baja de la distribución.

Otra condición propuesta por Dalton es la llamada **incrementos iguales de ingreso** e indica que si una distribución se obtiene de otra sumándole una constante **k**, entonces el índice disminuye cuando **k > 0** y aumenta en caso contrario.

3. **Simetría**. La condición de simetría impone a los índices que su valor no se altere cuando las mediciones de la desigualdad se hacen a un nivel donde los ingresos son iguales. Es decir, si dos individuos ubicados a un mismo nivel de la distribución intercambian sus ingresos, la medida de desigualdad debe mantenerse invariante e imparcial:

$$\mathbf{I(x_1, x_2) = I(x_2, x_1)} \qquad \mathbf{(1)}$$

4. **Condición de cambio relativo**. Supóngase que estamos interesados en conocer cómo se encuentra repartida la extensión agrícola de un país. Si se efectuara una redistribución de la tierra de un terrateniente en favor de un agricultor de autoconsumo, el coeficiente de concentración debiera reducirse en una proporción mayor que si la transferencia se generara de un agricultor de tamaño medio a un pequeño propietario.

Obsérvese que a diferencia de la condición de Pigou-Dalton, la propiedad de cambio relativo exige que exista una relación no lineal en el cambio experimentado por el indicador. Si un índice cumple con la condición de cambio relativo, automáticamente satisface el criterio de Pigou-Dalton; sin embargo, el razonamiento a la inversa no es válido.

5. **Independencia de Tamaño**. El índice de desigualdad debe proporcionar el mismo valor para dos poblaciones independientemente de su tamaño, siempre y cuando las proporciones de individuos para cada nivel de ingresos sea la misma, lo que implica que las curvas de Lorenz son iguales. Esta propiedad también se conoce como **independencia a la replicación de la población** y se enuncia como se presenta a continuación:

Si **m** grupos de **n** elementos cada uno tienen la misma distribución de probabilidad, entonces en una población de tamaño **n**x**m** la desigualdad deberá ser la misma que la que se obtiene de los grupos que la forman (ejemplos de esta propiedad pueden consultarse en Dasgupta, Sen y Starret, 1973).

6. **Consistencia con la Ordenación de la Curva de Lorenz (Dominancia Estocática)**. La curva de Lorenz es una manera alternativa de analizar la distribución del ingreso. Se dice que una curva domina a otra en el orden de Lorenz, si ésta se encuentra por encima de aquella en todos los puntos de la curva, salvo en los valores extremos. En este sentido, un índice será consistente con el orden de Lorenz cuando asume un valor menor para la distribución dominante con relación a la dominada. Los índices consistentes con este orden satisfacen el resto de las propiedades anteriores (Kakwani, 1980). Asimismo, un índice que es consistente con el orden de Lorenz debe ser convexo (Rothschild y Stiglitz, 1973).

7. **Decrecimiento de Efecto ante Transferencias**. Implica que las transferencias equivalentes entre individuos equidistantes tienen mayor efecto cuando ambos están ubicados en la parte baja de la distribución. Es decir, si y_1, y_2, y_3, y_4 satisfacen que $y_1 < y_2 < y_4$; $y_1 < y_3 < y_4$; $y_2 - y_1 = y_4 - y_3$, entonces una transferencia de y_2 a y_1 afectará más el valor del índice que si esta se efectuara de y_4 a y_3. Los indicadores estadísticos no satisfacen esta propiedad.

8. **Decrecimiento Relativo del Efecto ante Transferencias de Ingresos**. Es una propiedad más estricta que la anterior, ya que asigna mayor importancia relativa a las transferencias que se efectúan en la parte baja de la distribución, incluso en aquellos casos en que la diferencia de ingresos entre los dos individuos “más pobres” sea considerablemente menor que la de una pareja ubicada en la parte superior de la distribución. Si $y_1 < y_2 < y_4$; $y_1 < y_3 < y_4$; $y_2 / y_1 = y_4 / y_3$, entonces transferencias equivalentes desplazarán el índice hacia la igualdad si se realizan de y_2 a y_1, más que si fueran de y_4 a y_3, a pesar de que la distancia recorrida en el segundo caso sea mayor. Esta propiedad sólo la satisfacen algunos índices normativos.

10. **Descomposición Aditiva**. Un índice cumple con esta propiedad cuando puede calcularse para subgrupos, de tal forma que sea posible identificar la proporción de la desigualdad explicada por cada uno de ellos (Shorrocks, 1980). Supóngase una población de tamaño n con ingresos $y_1, y_2, \dots, y_n$ que se divide en k grupos e $y = (y_1, y_2, \dots, y_k, \dots, y_K)$, donde $y^k = (y^k_1, \dots, y^k_{nk})$. Entonces, se dice que un índice de desigualdad se puede descomponer en forma aditiva si existe un vector de ponderaciones $W^k = (w^1, \dots, w^k, \dots, w^K)$ que depende del vector de ingreso medio de la distribución $\mu = (\mu^1, \mu^2, \dots, \mu^k, \dots, \mu^K)$ y $n = (n^1, n^2, \dots, n^k, \dots, n^K)$, tal que se cumple

$$I(y) = I_w(y) + I_e(y) = \sum_{i=1}^{n} w^k(\mu,n)\, I(y^k) + I(\mu^1 e_{n1}, \mu^2 e_{n2}, \dots, \mu^3 e_{nk}) \qquad (2)$$

En donde I_w es una suma ponderada de la desigualdad dentro de cada grupo, y los ponderadores son factores que dependen del ingreso medio y del tamaño del grupo, y e_{n1}, en_2, ... , en_g son vectores unitarios de dimensión adecuada.

CUADRO 1
INFORMACIÓN NECESARIA PARA EL CÁLCULO DEL INDICE DE GINI

Decil	% de Hogares (p_i)	% de Ingreso (y_i)	% acumulado de hogares (P_i)	% acumulado de Ingreso (Y_i)
1	10	y_1	10	$Y_1=y_1$
2	10	y_2	20	$Y_2=Y_1+y_2$
3	10	y_3	30	$Y_3=Y_2+y_3$
4	10	y_4	40	$Y_4=Y_3+y_4$
5	10	y_5	50	$Y_5=Y_4+y_5$
6	10	y_6	60	$Y_6=Y_5+y_6$
7	10	y_7	70	$Y_7=Y_6+y_7$
8	10	y_8	80	$Y_8=Y_7+y_8$
9	10	y_9	90	$Y_9=Y_8+y_9$
10	10	y_{10}	100	$Y_{10}=Y_9+y_{10}$

iv) Calcular el índice de Gini conforme a alguna de las muchas expresiones que existen para el caso en que se trabaja con datos **agrupados**. Por ejemplo:

$$CG = 1 - \sum_{i=1}^{n} x_i(Y_i + Y_{i+1}) \qquad (7)$$

$$CG = \frac{1}{10000}\left[\sum_{i=1}^{n} (X_i Y_{i+1} - X_{i+1} Y_i)\right] \qquad (8)$$

$$CG = 1 - \sum_{i=1}^{n} (X_{i+1} - X_i)(Y_i + Y_{i+1}) \qquad (9)$$

en donde n representa el número de grupos, $\mathbf{x_i}$ la proporción (o porcentaje) de población en el grupo i; $\mathbf{X_i}$ la proporción acumulada de población en el grupo i; $\mathbf{Y_i}$ el ingreso acumulado en el grupo i.

IV.1 La Variable Apropiada para la Medición de la Desigualdad

En el trabajo empírico existe un debate sobre cuál es la variable que se considera más apropiada para evaluar la concentración del ingreso. De hecho, la principal controversia surge al momento de elegir entre el **ingreso total del hogar** o el **ingreso per cápita**.

Al respecto, es válido afirmar que ambas variables pueden ser adecuadas, dependiendo de los objetivos de la investigación que se quiera llevar a cabo. Por lo tanto, primero debiera especificarse cuál es la finalidad de la medición, para luego proceder a elegir en consecuencia la variable correspondiente.

En ciertos casos puede justificarse el uso del ingreso total familiar, por el simple hecho de que se considera que es el hogar la unidad de consumo en la cual se concentran las percepciones de ingreso de sus miembros y se decide sobre el destino de los recursos. Asimismo, el hogar es donde se comparten todos los bienes y servicios colectivos que son adquiridos con el presupuesto familiar, y el contexto que sirve para que sus miembros se formen y desarrollen para el proceso productivo. En ese sentido, cuando el propósito del investigador es analizar la distribución del ingreso entre los hogares, no debe existir duda en cuanto a que la variable que debe utilizarse es el ingreso corriente total del hogar[4].

Pero también puede darse un cuadro alternativo. En los últimos años, diversos investigadores han orientado el estudio del bienestar hacia las personas, y por tanto han insistido en la necesidad de que deben ser éstas el foco de análisis y de atención de la política social, a la vez que debiera ponerse mayor énfasis a las modalidades de asignación de recursos al interior de los hogares. En ese contexto, se reconoce que el tamaño de la familia es una variable de gran incidencia en la medición de la desigualdad y la pobreza, debido a que la constatación empírica confirma que los hogares de menores ingresos están integrados por un número mayor de personas, con relación a aquellos que se ubican en la parte alta de la distribución, lo cual tiene relación con mayores necesidades, pero también con las economías de escala que se generan en el uso de los bienes públicos al interior de los hogares.

De allí entonces que, si el objetivo de la investigación se centra en el estudio del bienestar individual, parece pertinente utilizar el ingreso per cápita del hogar como variable de referencia para evaluar la concentración del ingreso, ya que de lo contrario el ingreso familiar llevaría a subestimar el nivel de concentración del ingreso personal por el simple hecho de que se hace caso omiso de las necesidades que surgen debido a la presencia de un mayor número de personas.

Más allá de esto, sin embargo, y pese a que muchos estudios efectuados hacen uso de esta variable para medir la concentración del ingreso, es preciso señalar que en estricto rigor técnico el ingreso per cápita no sería tampoco la variable más apropiada para estudiar el bienestar individual, ya que en su cálculo no se considera el hecho de que las necesidades de los menores pueden diferir de las de los adultos, además de que existen en el hogar bienes públicos que dan servicio a todos sus miembros independientemente del tamaño de la familia. De modo que si el interés de los investigadores es mostrar realmente qué sucede con el nivel de vida de cada uno de los miembros del hogar, la manera correcta de proceder para evaluar la inequidad existente lleva a la necesidad de utilizar como factor de corrección del ingreso escalas de equivalencia que permitan construir una variable del **ingreso equivalente del hogar**.

Naturalmente, existen diferencias en los coeficientes de concentración del ingreso cuando el análisis de un mismo conjunto de datos se realiza a partir de las dos variantes mencionadas. Para ejemplificarlo, considérese la muestra de 30 hogares que se presenta en el cuadro 2.

---

4 Existen algunas definiciones más limitadas de ingreso que se captan en algunas de las encuestas de hogares vigentes en la región (monetario, remuneraciones al trabajo, etc.). Sin embargo, para ser consistente con el marco conceptual derivado del Sistema de Cuentas Nacionales (SCN), la mejor aproximación al presupuesto familiar se logra al considerar el ingreso corriente total del hogar (monetario más no monetario).

Con base en estos datos, si el interés es analizar la distribución familiar del ingreso, utilizando como variable de análisis el **ingreso total** del hogar, el valor del coeficiente de Gini es **0.8460**. Por el contrario, si el objetivo es conocer cómo se distribuye el ingreso personal, y se elige como variable de cálculo el **ingreso per cápita,** el nivel de concentración del ingreso aumenta a **0.8776**, lo que representa una diferencia relativa de 3.74% entre ambos valores.

CUADRO 2

Ingreso Total	Tamaño del Hogar	Ingreso Per cápita
25	7	3.57
29	7	4.14
38	6	6.33
49	6	8.17
50	5	10.00
128	5	25.60
155	7	22.14
159	7	22.71
258	4	64.50
369	4	92.25
423	4	105.75
536	3	178.67
569	5	113.80
639	4	159.75
698	6	116.33
719	7	102.71
789	4	197.25
1259	3	419.67
1278	4	319.50
1295	5	259.00
1459	4	364.75
1594	3	531.33
2587	4	646.75
2589	4	647.25
3574	3	1191.33
3697	3	1232.33
4225	5	845.00
12369	4	3092.25
15632	5	3126.40
69845	4	17461.25
127036	142	31370.50

IV.2 El Efecto del Número de Grupos en el Cálculo del Indice de Gini

Como premisa de análisis es necesario señalar que el cálculo de la concentración del ingreso a partir de datos agrupados introduce sesgos de subestimación. Esto se debe a la pérdida de información individual que se genera al no considerar la desigualdad dentro de cada grupo.

Desde un punto de vista teórico, la curva de Lorenz es un continuo de n puntos; por lo tanto, plantear el análisis de la distribución con datos agrupados consiste en aproximar un polinomio de grado n, a partir de un conjunto limitado de puntos $m < n$. De esta forma, en la medida que m se aproxime a n ($m \rightarrow n$), el error de estimación deberá reducirse. Por el contrario, cuando el número de intervalos formados sea muy pequeño ($m \rightarrow 0$), se incrementará en forma considerable el error de aproximación, y por tanto se subestimará de manera importante el verdadero nivel de desigualdad.

Lo anterior es fundamental tenerlo presente, ya que en la práctica es muy común que se calcule la concentración del ingreso a partir de particiones que consideran el 20% de las observaciones, denominadas quintiles, sin hacer conciencia de que se trata de una aproximación muy gruesa, que seguramente subestima en magnitud importante el nivel de desigualdad.

En el cuadro 3 se presentan valores del coeficiente de Gini calculados para diferentes agrupaciones. Si se define como parámetro de referencia el valor del coeficiente a partir de la formación de deciles, se observa que en la medida que el número de grupos disminuye, el coeficiente de concentración del ingreso de esta población hipotética tiende también a disminuir, mostrando una diferencia absoluta de 0.1568 puntos en el caso en que la desigualdad se evalúa a partir de la formación de cinco grupos.

CUADRO 3
VARIACIONES EN EL INDICE DE GINI SEGÚN EL NÚMERO DE GRUPOS

Número de Grupos	Coeficiente de Gini	Diferencias con relación al valor con 10 grupos	% de diferencia
10	0.3709	---	---
9	0.3686	-0.0023	-0.60
8	0.3588	-0.0121	-3.26
7	0.3224	-0.0485	-13.08
6	0.2778	-0.0931	-25.10
5	0.2141	-0.1568	-42.28
4	0.1577	-0.2132	-57.48
3	0.1078	-0.2631	-70.94
2	0.0488	-0.3221	-86.84

Otro aspecto que debe señalarse son las diferencias que existen en el valor del indicador cuando el cálculo se realiza con los datos individuales, con relación a lo que resulta cuando se forman grupos de observaciones, tal como lo muestra el cuadro 4.

CUADRO 4
COEFICIENTE DE DESIGUALDAD DE GINI PARA DATOS DESAGREGADOS Y AGRUPADOS

Variable de Análisis	Datos Desagregados	Datos Agregados
Ingreso Total	0.8460	0.7861
Ingreso Per Cápita	0.8776	0.8127

Utilizando los datos del cuadro 4 se observa que para cualquiera de las dos variantes de análisis, el coeficiente de Gini calculado con datos agregados genera una subestimación de casi 8% en el verdadero valor de la desigualdad, correspondiente al cálculo con datos desagregados.

IV.3 Variables para el Ordenamiento de los Hogares y el Cálculo del Coeficiente de Gini

La manera convencional de medir la concentración en un conjunto de datos consiste en generar, primeramente, un ordenamiento de las observaciones en función de la variable que se desea analizar. Por ejemplo, si se busca conocer el nivel de concentración del ingreso entre las familias, se procede como sigue:

i) Ordenar los hogares conforme a su ingreso (total o per cápita);

ii) Definir el número de intervalos a formar;

iii) Calcular para cada grupo el porcentaje de observaciones y del ingreso que concentra;

iv) Aplicar alguna fórmula de cálculo para datos agrupados que permita conocer la distribución de la variable en la población objeto de estudio.

En el trabajo práctico surgen diferentes opciones para el ordenamiento de los hogares, así como para la elección de la variable a utilizar en el análisis. A continuación se presentan algunas de las alternativas para el ordenamiento de los hogares, de las variables a utilizar para efectuar los cálculos, y de las combinaciones entre ambas. A fin de facilitar la presentación, en adelante se supondrá que lo que interesa es estudiar la distribución del ingreso entre los hogares.

Opciones:

a) Ordenar a los hogares según su ingreso total (**Y**) y acumular el ingreso total del hogar como variable de análisis (**Y**).

b) Ordenar a los hogares según su ingreso per cápita ($\mathbf{Y_p}$) y acumular el ingreso per cápita ($\mathbf{Y_p}$).

c) Ordenar los hogares según su ingreso per cápita ($\mathbf{Y_p}$) y acumular el ingreso total del hogar (**Y**).

d) Ordenar los hogares según su ingreso total (**Y**) y acumular el ingreso per cápita ($\mathbf{Y_p}$).

En las dos primeras opciones, la variable que genera el ordenamiento coincide con aquella que se utiliza para efectuar el análisis. En cambio, en las alternativas c) y d) existen diferencias entre la variable del ordenamiento y aquella que se emplea en el cálculo del coeficiente de desigualdad.

Al respecto, debe hacerse notar que las alternativas c) y d) pueden provocar inconsistencias en el análisis que conduzcan a la subestimación del verdadero nivel de la desigualdad. De hecho, cuando a partir de una combinación de las variables como las señaladas se calcula el coeficiente de Gini, el indicador se conoce comúnmente como **pseudo-Gini**, ya que la variable que ordena no coincide con la que se utiliza para calcular la desigualdad, situación que difiere del planteamiento teórico que da sustento a la lógica de construcción del índice de Gini[5].

Normalmente los hogares pobres tienden a estar conformados por un mayor número de miembros, por lo que, bajo este procedimiento, puede suceder que algunos de los grupos formados representen un porcentaje de población mayor que el que le corresponde según el ordenamiento definido, como se aprecia en el ejemplo hipotético del cuadro 5. Allí se muestra que, pese a que el primer decil agrupa el 10% de los hogares, el mismo concentra el 14.1% de las personas. En circunstancias que, para ser consistente con el criterio utilizado para el cálculo del índice de Gini, en una situación igualitaria (CG=0) le debiera corresponder el 14.1% del ingreso total, y no el 10% como lo supone el ordenamiento llevado a cabo.

---

5 Como se verá más adelante, el hecho de que puedan existir diferencias entre la variable que ordena y la distribución al interior de los grupos formados, conduce a que este índice no se pueda descomponer en forma aditiva.

CUADRO 5
DECILES DE HOGARES ORDENADOS SEGÚN SU INGRESO PER CÁPITA

% Hogares	%person.	% ing. Total	%ing. Acumulado	Ing. Total	ing. Medio	personas	ing. Per/cáp.	tam. hogar
10	14.08	0.07	0.07	92	9.2	20	4.60	2.0
10	12.68	0.20	0.27	254	25.4	18	14.11	1.8
10	11.27	0.43	0.70	545	54.5	16	34.06	1.6
10	10.56	1.19	1.89	1511	151.1	15	100.73	1.5
10	10.56	1.50	3.39	1906	190.6	15	127.07	1.5
10	9.15	2.06	5.45	2620	262.0	13	201.53	1.3
10	7.75	3.15	8.60	3996	399.6	11	363.27	1.1
10	7.75	5.33	13.93	6770	677.0	11	615.45	1.1
10	7.75	9.05	22.98	11496	1149.6	11	1045.09	1.1
10	9.15	77.02	100.00	97846	9784.6	13	7526.61	1.3
Coeficiente de Gini		0.7854						

Supóngase que se dispone de una muestra de 30 hogares con información sobre el ingreso total y el número de personas que integran cada uno de ellos, y se desea conocer el nivel de concentración del ingreso por medio del coeficiente de Gini. En esas condiciones, se requiere efectuar el ordenamiento de los hogares y seleccionar la variable apropiada para el análisis siguiendo alguna de las opciones señaladas más arriba.

Como se indicó, la opción a) consiste en ordenar los hogares según su ingreso total, y esta misma variable es la que se utiliza para calcular el coeficiente de Gini. En este caso, se obtiene un valor para la desigualdad del ingreso entre los hogares de 0.7861, tal como se registra en el cuadro 6.

CUADRO 6
INDICE DE CONCENTRACIÓN DEL INGRESO DE LOS HOGARES
OPCIÓN A): DECILES DE HOGARES ORDENADOS POR INGRESO TOTAL Y SE ACUMULA EL INGRESO TOTAL

Decil	% Hogares	%person	% ing. Total	%ing. Acumulado	ing. Total	ing. Medio	Personas	ing. Per/cáp.	tam. hogar
1	10	14.08	0.07	0.07	92	9.2	20	4.60	2.0
2	10	11.27	0.18	0.25	227	22.7	16	14.19	1.6
3	10	12.68	0.45	0.70	572	57.2	18	31.78	1.8
4	10	7.75	1.05	1.75	1328	132.8	11	120.73	1.1
5	10	10.56	1.50	3.25	1906	190.6	15	127.07	1.5
6	10	9.86	2.18	5.43	2767	276.7	14	197.64	1.4
7	10	9.15	3.17	8.60	4032	403.2	13	310.15	1.3
8	10	7.75	5.33	13.93	6770	677.0	11	645.45	1.1
9	10	7.75	9.05	22.98	11496	1149.6	11	1045.09	1.1
10	10	9.15	77.02	100.00	97846	9784.6	13	7526.62	1.3
Coeficiente de Gini			0.7861						

A su vez, la información de ese mismo cuadro permite inferir que los hogares ubicados en los primeros deciles están compuestos por un mayor número de personas. De modo que si el objetivo es el estudio de la concentración del ingreso entre los individuos, el análisis debiera privilegiar el uso de una variable que tenga en cuenta esas diferencias en el tamaño de la familia, por lo que debiera optarse por el ingreso per cápita tanto para el ordenamiento de las observaciones como para el cálculo del nivel de desigualdad. Los resultados de la aplicación de este procedimiento se presentan en el cuadro 7.

CUADRO 7
INDICE DE CONCENTRACIÓN DE LOS INGRESOS INDIVIDUALES
Opción b): Deciles de personas ordenadas por su ingreso per cápita y se acumula el ingreso per cápita

Decil	% personas	% ing. Per/cápita	%ing Acumulado	ing. Total	ing. Per/cáp.	personas
1	10	0.04	0.04	54.0	3.86	14
2	10	0.08	0.12	107.0	7.64	14
3	10	0.22	0.34	275.9	19.70	14
4	10	0.50	0.84	638.6	45.62	14
5	10	1.13	1.97	1440.3	102.88	14
6	10	1.41	3.38	1792.2	128.01	14
7	10	2.57	5.95	3259.0	232.79	14
8	10	4.92	10.87	6244.5	446.04	14
9	10	9.20	20.07	11681.5	834.39	14
10	10	79.93	100.00	104543.0	6346.44	16
Coeficiente de Gini		**0.8127**				**142**

Como se señaló anteriormente, cuando se procede bajo esta opción es lógico esperar que el coeficiente de concentración del ingreso arroje un valor más elevado. En efecto, mientras en el primer caso el coeficiente de Gini de los hogares es de 0.7861, en el segundo, la concentración del ingreso entre los individuos es de 0.8127.

Los resultados de los cuadros 6 y 7 expresan la manera apropiada de analizar la concentración del ingreso familiar e individual, respectivamente. Sin embargo, a veces se combinan el ingreso total del hogar y el ingreso per cápita para calcular la desigualdad, tal como se muestra en los cuadros 8 y 9.

En el cuadro 8 el ordenamiento de los hogares se efectúa según su ingreso per cápita, en tanto que se utiliza el ingreso total del hogar para calcular el índice de Gini. Bajo esta opción se pierde el efecto de la agrupación por el ingreso per cápita, ya que el ingreso total no cumple la condición básica de estar ordenado en forma creciente ($\mathbf{y_1 \leq y_2 \leq \ldots \leq y_n}$), y por ende se subestima el verdadero nivel de la desigualdad.

CUADRO 8

Opción c): Deciles de hogares ordenados por su ingreso per cápita y se acumula el ingreso total

Decil	% Hogares	%person.	% ing. Total	%ing Acumulado	ing. Total	ing. Medio	personas	ing. Per/cáp.	tam. hogar
1	10	14.08	0.07	0.07	92	9.2	20	4.60	2.0
2	10	12.68	0.20	0.27	254	25.4	18	14.11	1.8
3	10	11.27	0.43	0.70	545	54.5	16	34.06	1.6
4	10	10.56	1.19	1.89	1511	151.1	15	100.73	1.5
5	10	10.56	1.50	3.39	1906	190.6	15	127.07	1.5
6	10	9.15	2.06	5.45	2620	262.0	13	201.53	1.3
7	10	7.75	3.15	8.60	3996	399.6	11	363.27	1.1
8	10	7.75	5.33	13.93	6770	677.0	11	615.45	1.1
9	10	7.75	9.05	22.98	11496	1149.6	11	1045.09	1.1
10	10	9.15	77.02	100.00	97846	9784.6	13	7526.61	1.3
Coeficiente de Gini				0.7854					

Nuevamente se aprecia en el cuadro 8 que el porcentaje de personas agrupadas en el primer decil (14.1%) supera el valor máximo que admite la partición teórica realizada (10%), y por lo tanto viola el supuesto básico para la elaboración de la curva de Lorenz y el cálculo del coeficiente de Gini, en virtud de que los hogares con mayor número de personas se ubican en los primeros deciles, generando una contradicción entre el porcentaje de ingreso que debieran retener en condiciones de igualdad perfecta y el que les asigna el procedimiento de cálculo aplicado.

Este procedimiento conduce a que se genere una subestimación en la concentración del ingreso, tal como se observa en el cuadro 8. En efecto, mientras que a partir de los cuadros 6 y 7 se observaron coeficientes de desigualdad de 0.7861 y 0.8127, la opción c) derivó un índice de 0.7854.

CUADRO 9

Opción d): Deciles de hogares ordenados por su ingreso total y se acumula el ingreso per cápita

Decil	% Hogares	% Personas	% ing. Total	%ing Acumulado	Ing. Medio	Total de personas	ing. Per/cáp.	tam. hogar
1	10	14.08	0.04	0.04	14.05	20	0.70	2.0
2	10	12.68	0.13	0.17	40.31	18	2.23	1.8
3	10	11.27	0.36	0.53	112.81	16	7.06	1.6
4	10	10.56	0.96	1.49	300.71	15	20.04	1.5
5	10	10.56	1.24	2.73	389.88	15	25.99	1.5
6	10	8.45	2.02	4.75	634.92	12	52.91	1.2
7	10	7.75	3.52	8.27	1103.92	11	100.36	1.1
8	10	7.75	5.82	14.09	1825.33	11	165.94	1.1
9	10	7.75	10.42	24.52	3268.67	11	297.15	1.1
10	10	9.15	75.48	100.00	23679.9	13	1821.53	1.3
Coeficiente de Gini				**0.7867**				

Finalmente, el cuadro 9 presenta la alternativa en que los hogares se ordenan según su ingreso total, y el nivel de la desigualdad se determina a partir del ingreso per cápita. En este caso, el valor del índice de Gini es ligeramente mayor al que se presentó en el cuadro 6, pero resulta inferior al mostrado en el cuadro 7 y casi igual al registrado en el cuadro 8.

Recuadro 5

Las Comparaciones Internacionales de la Desigualdad

Cada vez es más común que se realicen estudios que tengan como objetivo comparar el nivel de bienestar entre un conjunto de países, y generar ordenamientos que permitan conocer en cuáles de ellos existe menor desigualdad.

Ante una situación como la descrita, se acude a diversas fuentes de información en donde se concentran datos agregados, y las menos de las veces se tiene acceso a los microdatos originales y se efectúan procesamientos específicos que garanticen la estandarización de los procedimientos de cálculo y la identificación de las variables apropiadas.

A pesar de que para todos resulta interesante efectuar comparaciones internacionales, es muy frecuente que en aras de disponer de una amplia base informativa, en los análisis se omita hacer referencia a los procedimientos operativos que están detrás de los indicadores comparados y se actúa como si todas las estimaciones estuvieran generadas con la misma calidad estadística y los métodos de cálculo de la desigualdad provinieran de un algoritmo estándar que garantiza la perfecta comparabilidad entre el conjunto de observaciones estudiadas.

De hecho, son casi inexistentes los estudios que se ocupan de evaluar la varianza de muestreo de las estimaciones sobre ingresos, y aquellos que comparan los resultados obtenidos con una fuente de información exógena que permita validar la robustez estadística de la información generada por las encuestas. Además, tampoco es práctica común que se hagan análisis sobre la manera en que se agruparon los hogares, el concepto de ingreso que se utilizó para su ordenamiento, el número de grupos formados, ni tampoco se hace referencia a la fórmula que se aplicó para medir el nivel de la desigualdad, y no se cuestiona si los índices provienen de datos agrupados o corresponden a cálculos realizados con observaciones desagregadas.

Todas estas variantes tienen influencia sobre el nivel del coeficiente de concentración del ingreso, y en particular sobre los valores resultantes del índice de Gini. De esta manera, un análisis efectuado con la mejor voluntad puede derivar en un ejercicio que no haga justicia a ninguno de los países comparados, y por tanto se generen juicios de valor que deriven en confusión debido a que, en el fondo, el conjunto de datos recolectado no admite comparabilidad alguna.

Es muy importante enfatizar la necesidad de cuidar todos los aspectos asociados con la medición: el objetivo de la comparación; la unidad de análisis estudiada (hogares o personas); la temporalidad de los datos; la cobertura geográfica de las encuestas; el marco metodológico para el concepto de ingreso analizado; la variable que utilizó para ordenar y agrupar a los hogares; la que se aplicó para acumular el ingreso; el número de grupos formados; la fórmula utilizada en el cálculo; si se utilizaron microdatos o se trata de una estimación con información agregada. Estos son algunos de los aspectos que influyen sobre el valor de los indicadores de desigualdad y en particular sobre el coeficiente de Gini.

No puede efectuarse de manera acrítica un estudio internacional, sin tener certeza de que los procedimiento de estandarización de los datos garantizan la comparabilidad absoluta de los niveles de desigualdad en los países.

El interés por efectuar la comparación, debe necesariamente acompañarse de la preocupación por identificar un estándar común. En los países europeos que participan en el proyecto “Luxemburg Income Study (LIS)”, se acostumbra evaluar la desigualdad utilizando una medida estándar de ingreso, la cual se deflacta por una escala de equivalencia común, para garantizar que los niveles de concentración del ingreso sean estrictamente comparables.

Esta situación conlleva el que los indicadores de desigualdad de estos países no puedan compararse en forma directa con el resto del mundo, a menos que se logre operacionalizar una definición de ingreso compatible y los recursos financieros de los hogares se ajusten por la misma escala de equivalencia.

En el caso de América Latina y el Caribe la situación es muy diversa. La cobertura de las encuestas no necesariamente es compatible entre países; la temporalidad y la naturaleza de los estudios también exhiben diferencias importantes; el concepto de ingreso relevado no siempre es el mismo; en algunos casos se dispone de información sobre encuestas de ingresos y gastos y la mayoría de las veces se trabaja con los ingresos reportados en las encuestas de empleo.

Todas estas situaciones representan elementos que deben considerarse cuando se decide comparar el nivel de la concentración del ingreso entre países.

Un elemento adicional muy importante es la necesidad de involucrar los aspectos demográficos que están detrás de la medición y que se relacionan con el tamaño y composición de las familias, y las economías de escala que se generan al interior de los hogares.

La dinámica demográfica en los países de la región es altamente diferenciada, lo cual induce a que cualquier tipo de comparación que se establezca y en donde se involucre el tamaño del hogar, debe considerar que el número de miembros y la composición de un hogar de un país sudamericano, es muy diferente de aquella que se observa en la mayoría de los países de América Central, en donde las altas tasas de fecundidad continúan influenciando el crecimiento natural de la población.

Lo anterior permite afirmar que las comparaciones regionales deben estar basadas en una noción de ingreso que involucre en la medición las economías de escala que se generan por la alta presencia de niños en los hogares en condiciones de pobreza, así como las equivalencias.

V. LA DESAGREGACIÓN DEL INDICE DE GINI

La descomposición juega un papel central en el análisis de la desigualdad, ya que es importante conocer de dónde provienen los mayores desequilibrios que afectan la equidad entre los hogares.

Es necesario partir estableciendo la diferencia que existe entre la descomposición que se genera por las distintas fuentes de ingreso de las familias, y aquella que se efectúa a partir del estudio de las características del hogar o de algunos de sus miembros. En esta distinción se debe reconocer la existencia de factores que inciden sobre las percepciones que reciben los individuos (edad, sexo, nivel de escolaridad, ocupación, experiencia), mientras que otras afectan directamente las fuentes de ingreso (inflación, negociaciones salariales, etc.), y también se presentan situaciones que afectan tanto a los ingresos como a los individuos (variaciones negativas del indicador de actividad económica).

A pesar de que el índice de Gini es ampliamente utilizado en los estudios sobre la desigualdad, es necesario reconocer que este indicador le concede mayor peso a las observaciones que se ubican en la parte central de la distribución, y no obstante su popularidad y fácil procedimiento de cálculo, no cumple con la propiedad de descomposición aditiva, excepto para aquellas situaciones en que los diversos grupos se ordenen por niveles de ingresos. Estos significa que el coeficiente de Gini, calculado para un determinado número de subgrupos, no coincide con el valor del coeficiente estimado a partir de un ordenamiento de la población total por niveles de ingreso.

V.1 Descomposición de la Desigualdad por Subgrupos de Población

En este contexto de análisis, el objetivo central es determinar cuál es el efecto en el nivel general de la desigualdad del agrupamiento de la observaciones, analizando de qué manera la concentración del ingreso puede ser explicada por las diferencias observadas entre los grupos en que la población ha sido dividida.

Es frecuente que este tipo de consideraciones se efectúe utilizando variables asociadas al contexto geográfico de residencia del hogar, el nivel de escolaridad de las personas, el tipo de ocupación de las personas activas y, en ocasiones, resulta relevante involucrar el género de las personas.

La base teórica de las propuestas que existen para descomponer los índices de desigualdad surge de los conceptos estadísticos del análisis de varianza. Desde esta perspectiva, la idea de esta metodología consiste en determinar qué porcentaje de la variabilidad está explicada por cada uno de los factores que inciden en su comportamiento. En el caso del estudio de la desigualdad, el objetivo se orienta a determinar qué porcentaje de la variabilidad en la distribución del ingreso está explicado por factores asociados a la escolaridad, el mercado de trabajo, a la riqueza acumulada, al patrimonio, etc.

En términos estadísticos, la varianza de una variable como el ingreso (Y), asociada a una agrupación determinada y dado un conjunto de factores sociodemográficos (x_1, x_2, ..., x_n), se puede expresar como se muestra a continuación:

$$\mathbf{V(Y/\, x_1,x_2,\ldots\, x_n) = W_I\, V(Y_1) + W_E\, V(Y_2) = \text{dentro de los grupos} + \text{entre grupos}} \qquad \mathbf{(10)}$$

En donde W_I y W_E representan la importancia relativa de cada grupo en la población, y Y_1 e Y_2, son los ingresos retenidos por los grupos respectivos.

Cuando se habla de la varianza “entre grupos”, se hace referencia a la variabilidad que se tendría si se analiza el ingreso de cada grupo representado por el valor medio de cada intervalo. Por su parte, la connotación “dentro de los grupos” se determina a partir del promedio ponderado de las varianzas dentro de cada grupo, en donde los ponderadores se calculan considerando la importancia relativa de cada grupo en la población y del ingreso medio.

Al calcular el cociente de la variación “entre grupos” entre la varianza total, se obtiene la variabilidad generada por el ingreso, mientras que el otro caso se interpreta como la desigualdad que no es generada por la clasificación elegida.

A pesar de que la descomposición aditiva de los índices de desigualdad es una propiedad ampliamente aceptada, algunos autores (Sen, 1992) han cuestionado el hecho de que cuando un agrupamiento presenta algún tipo de interdependencia, no es posible efectuar una descomposición exacta y por tanto debe considerarse la existencia de un término residual.

Se han efectuado diferentes investigaciones en donde se han propuesto condiciones que deben cumplir los índices para que se puedan descomponer en forma aditiva (Shorrocks, 1980). Este autor dice que para que una medida se pueda descomponer en forma aditiva se debe poder expresar de la manera siguiente:

$$\mathbf{I(x) = I_{dentro} + I_{entre} = \sum w_g\, I(x_g) + I_{entre}} \qquad \mathbf{(11)}$$

En donde $\mathbf{x_1,x_2, \ldots , x_g}$ es una partición de la variable **x** en **g** subgrupos exclusivos y excluyentes, $\mathbf{I(x_g)}$ el índice de desigualdad dentro de cada subgrupos, $\mathbf{w_g}$ los ponderadores y el término **“entre”** está influenciado por las medias y los tamaños de la población.

En el caso de que las medidas se puedan descomponer en forma aditiva y además satisfagan los principios de transferencias, de población y de homogeneidad de ingresos, se obtiene la familia de índices de entropía generalizada, cuyos ponderadores se expresan de la manera siguiente:

$$\mathbf{W_g = [n_g / n]\, [\mu_g / \mu]\, \theta} \qquad \mathbf{(12)}$$

en donde n_g representa el tamaño del grupo g, μ_g la media, θ el parámetro de aversión a la desigualdad, y n y μ el tamaño y la media de la población, respectivamente.

En un trabajo posterior, Shorrocks (1984) propuso relajar las restricciones impuestas y sólo considerar la condición de que la inequidad global pueda expresarse a partir del tamaño de la población, la media de los ingresos y el índice de desigualdad de cada subgrupo de población. A pesar de esta situación, se afirma que el índice de Gini no puede ser expresado como un índice que se descompone en forma aditiva.

Una manera de analizar la descomposición del coeficiente de Gini fue propuesta por Foster y Sen (1997) y se presenta a continuación:

$$\mathbf{G(Y/\ x_1,\ x_2,\ ...,x_n) = variabilidad\ dentro\ de\ los\ grupos\ +\ variabilidad\ entre\ grupos + residuo} \quad \mathbf{(13)}$$

en donde la desigualdad "dentro", se obtiene como un promedio ponderado de los índices dentro de los grupos con ponderadores

$$\mathbf{W_g = [n_g\ /\ n]^2\ [\mu_g\ /\ \mu]} \qquad \mathbf{(14)}$$

n = tamaño de la población, n_g = tamaño del grupo g, μ_g = media del ingreso en el grupo g, μ = ingreso medio de la población, "entre" = valor del coeficiente de Gini calculado sobre los ingresos medios de cada grupo, "residuo" = valor residual no negativo

V.2 Descomposición de la Desigualdad por Fuentes de Ingresos

El objetivo del análisis por fuentes es determinar la contribución de cada concepto de ingresos a la desigualdad observada en la población de referencia.

Los métodos analizados en el apartado anterior no permiten efectuar dicha descomposición, debido a que frecuentemente las familias perciban ingreso de diversas fuentes, lo cual incide en que no se puedan formar agrupaciones mutuamente excluyentes.

La literatura consigna diferentes propuestas metodológicas para lograr el objetivo, pero es en Shorrocks (1982) en donde se denomina regla de descomposición a cualquier función que asigna valores adecuados a las participaciones relativas de cada factor, con la propiedad de que la suma de los ponderadores sea igual a la unidad.

Utilizando la teoría del análisis de varianza, la descomposición de la desigualdad generada por los ingresos se puede expresar como se muestra a continuación:

$$\sigma^2(Y) = \sum_k \sigma^2(Y^k) + \sum_{j\neq k} \sum_k \rho_{jk}\sigma(Y^j)\sigma(Y^k) \qquad \mathbf{(15)}$$

donde ρ_{jk} representa el coeficiente de correlación entre las fuentes de ingreso j y k.

En caso que exista correlación entre las distintas fuentes de ingreso, deberán considerarse en la ecuación (5) tantos efectos de interacción como correlaciones distintas de cero entre las fuentes de ingreso analizadas.

En el trabajo de Shorrocks (1982) se plantea el objetivo de encontrar una regla única de descomposición para determinar la influencia de cada fuente de ingreso a la desigualdad total. Así, se pide que la contribución a la inequidad no varíe en función del indicador elegido; además, se exige que cuando el ingreso de una o más fuentes se distribuya en forma equitativa esta situación no debe tener influencia en la desigualdad. Finalmente se pide que si dos componentes del ingreso tienen la misma importancia, su contribución a la desigualdad sea la misma.

El índice de Gini no satisface esta regla de descomposición. Sin embargo, Kakwani (1994) justifica la importancia de analizar su descomposición por fuente de ingreso, y por tanto sugiere la siguiente expresión:

$$G(Y) = \frac{2}{n^2\mu}\sum_i \left(i - \frac{n+1}{2} \right) \sum_k Y_i^k = \sum_k \frac{\mu_k}{\mu} \overline{G}(Y^k) \qquad \textbf{(16)}$$

μ_k representa el valor promedio de la k-ésima fuente de ingreso y

$$\overline{G}(Y^k) = \frac{2}{n^2\mu}\sum_i \left(i - \frac{n+1}{2} \right) Y_i^k \qquad \textbf{(17)}$$

la expresión anterior se conoce comúnmente como pseudo-Gini para el k-ésimo componente de ingreso, debido a que el ordenamiento de los ingresos al interior del grupo no coincide necesariamente con la que el que se observa al para la fuente de ingresos k, lo cual permite que $\overline{G}$ (Y^k) asuma valores negativos. Además, en el caso de que su nivel sea mayor al del índice global, la interpretación sugiere que en ese grupo se presenta una mayor desigualdad que en la población de referencia.

Por otra parte, para explicar el cambio en la desigualdad de las distintas fuentes de ingreso entre dos mediciones, Kakwani (*op cit*.) define la siguiente expresión;

$$\mathbf{G^* - G = \Sigma^k_{i=1} (w_k^* c_k^* - w_k c_k) = \quad D} \qquad \textbf{(18)}$$

$\mathbf{w_j = \mu_j / \mu}$, representa la importancia relativa de la j-ésima fuente de ingreso en el ingreso total, D se interpreta como la contribución del componente de ingreso k a la desigualdad total y c_k es el pseudo-Gini calculado para la k-ésima fuente de ingreso.

El valor de D a su vez se representa por medio de:

$$\mathbf{D = ½ ((w_k^* - w_k) (c_k + c_k^*) + ½ (c^*_k - c_k)(w_k - w^*_k)} \qquad \textbf{(19)}$$

VI. LAS FUENTES DE DATOS PARA LA MEDICIÓN DE LA DESIGUALDAD

En los libros de texto, así como en la mayoría de los artículos que hacen referencia a la medición de la desigualdad, las fórmulas que se proponen asumen que los datos sobre ingresos que se utilizan provienen de una muestra aleatoria $\mathbf{y_1, y_2, \ldots , y_n}$ de n hogares. Sin embargo, el trabajo empírico se efectúa a partir de datos provenientes de encuestas de hogares, en donde la identificación de las unidades de observación se realiza a partir de dos o más etapas de selección y la mayoría de las veces los hogares se eligen con probabilidades desiguales. Esta situación impone al menos la solución de dos problemas fundamentales asociados con la medición de la desigualdad: la ponderación de las observaciones y, dado que los coeficientes de desigualdad son valores estimados, conocer la precisión de los estimadores.

El hecho que cada una de las observaciones que se utiliza tenga asociada una probabilidad de selección conocida y distinta de cero, implica que se dispone de **n** parejas de datos $\mathbf{(y_1,f_1), (y_2,f_2), \ldots, (y_n, f_n)}$, donde f_i representa el inverso de la probabilidad de selección de la observación (p_i) que comúnmente se conoce como factor de expansión o ponderador.

Es evidente que cuando las observaciones tienen asociado un factor de expansión distinto, la medida de desigualdad se ve afectada por los valores de los ponderadores, y por lo tanto se generan diferencias entre los niveles de desigualdad calculados con observaciones muestrales (sin ponderar) o con los datos expandidos. Así, carece de sentido que se presenten indicadores de desigualdad calculados con datos muestrales, a sabiendas de que las probabilidades de selección de las observaciones difieren y esta situación incide sobre el valor del coeficiente de concentración del ingreso.

Por otra parte, es necesario recordar que cualquier indicador que se calcule a partir de los datos de una muestra tiene asociado un error de estimación, el cual se debe conocer a fin de determinar si la información generada tiene la precisión estadística requerida.

En la literatura económica este problema frecuentemente se pasa por alto, y prácticamente no existen estudios sobre la desigualdad económica que se acompañen de información que haga referencia a la precisión estadística del índice de Gini, o de cualquier otro indicador que de cuenta del nivel de desigualdad entre el grupo de hogares analizados. En la actualidad, este tema se ha comenzado a discutir, y se reconoce la necesidad de incorporar elementos que informen sobre la calidad estadística de los indicadores, al tiempo que se está consciente de que el tratamiento para datos agrupados es más complicado que cuando se trabaja con observaciones individuales. Por lo tanto, la mayoría de las investigaciones que atienden este problema tratan de aproximar el error estándar de los coeficientes de desigualdad a partir de las propiedades asintóticas de distribuciones muestrales, y por ende la aplicación de los resultados teóricos que se han derivado está condicionada a la disponibilidad de información para muestras suficientemente grandes en las que el supuesto de normalidad de las distribuciones se satisfaga (véase el Recuadro 3).

En el contexto de la inferencia estadística, las medidas de desigualdad se pueden clasificar de la siguiente manera: i) Indices que se pueden escribir en función de los valores medios de la variable monetaria que se está analizando. A este grupo pertenecen el índice de Theil, Atkinson y el coeficiente de variación al cuadrado. ii) Por otra parte, existe un conjunto de medidas que requieren un ordenamiento previo de los datos, como la curva de Lorenz y el coeficiente de Gini. En este caso, la observaciones individuales se analizan bajo el supuesto de que representan estadísticas de orden de la distribución.

En el caso de los índices que pertenecen a la clase i), la aplicación del teorema del límite central permite derivar aproximaciones asintóticamente normales al error estándar de la distribución, como las que se presentan en el Recuadro 3.[6]

Para el caso en que se trabaja con datos agrupados, la estimación de la varianza se puede aproximar por alguno de los siguientes métodos: muestras replicadas, el método de "jackknife" o aplicando la técnica del "bootstrap"[7].

Es importante enfatizar que los métodos que se utilizan para estimar la varianza de muestreo, y que se basan en las propiedades asintóticas de las distribuciones para muestran grandes, tienden a subestimar el verdadero error de muestreo de las estimaciones[8], por lo que, siempre que sea posible, se recomienda emplear un algoritmo de cálculo que considere las diferentes etapas que se involucran en la selección de la muestra, así como sus probabilidades de selección.

VII. CONCLUSIONES

La medición de la desigualdad es un tema que permanentemente ha estado presente en el debate económico. Es común que su análisis y evolución estén siempre presentes cuando se evalúa la gestión económica de los gobiernos y sus repercusiones sobre el nivel de vida de las familias, así como la

6 Para conocer los detalles técnicos se puede consultar Kakwani, N. (1990), "Large Sample Distribution of Several Inequality Measures with Applications to Cote d'Ivoire", LSMS Working Paper # 61, World Bank.

7 El procedimiento estadístico para la aplicación del "jackknife" y del "bootstrap" se puede consultar en Efron, B. y Tibshirani, R. (1986). "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Statistical Accuracy", *Statistical Science*, vol. 1, No. 1, pp.54-77.

8 Para conocer un análisis sobre la comparación entre este tipo de aproximaciones y las estimaciones basadas en el diseño de muestra puede consultarse Medina, F. (1998), "Los Errores de Muestreo en Encuestas Complejas", pp. 315-348 en *Memoria del Primer Taller Regional del MECOVI*, Aguascalientes, México. CEPAL, 1998.

equidad existente en una sociedad. Las formas de abordar la discusión son diversas y los indicadores utilizados varían conforme al propósito de la investigación y el énfasis que se le quiera dar a la magnitud del problema.

Además de la gran utilidad que representa utilizar el índice de Gini como una medida de concentración del ingreso, ha quedado en evidencia que existe una gran cantidad de indicadores que dan cuenta del grado de desigualdad existente en una población, y que cada uno de ellos responde a una base conceptual distinta, a un procedimiento de cálculo específico y se concentra en analizar parcialmente un aspecto particular del fenómeno de estudio.

Es preciso señalar que, los diferentes índices que han sido presentados tienen un carácter esencialmente ordinal, y por tanto los valores que generan no son comparables entre sí, a pesar de que sus cálculos se efectúen con el mismo conjunto de datos. Esto significa que carece de sentido comparar el valor del coeficiente de desigualdad de Gini con el del índice de concentración de Theil o de Atkinson.

Cuando se tienen distintas mediciones en el tiempo de un mismo indicador para una población de referencia, los valores son comparables sólo en un sentido ordinal, lo que permite concluir en qué situación existe menor desigualdad.

Otro aspecto que también se debe considerar en el trabajo empírico es que cada uno de los indicadores da cuenta de un aspecto específico de la desigualdad, y por tanto pueden generar ordenamientos diferentes.

Esta situación es de vital importancia cuando se analiza la desigualdad a partir de la comparación de dos o más curvas de Lorenz. Se pueden generar conclusiones contundentes sólo cuando una de ellas queda completamente contenida en la otra, en cuyo caso se dirá que alguna de las distribuciones analizadas domina en el orden de Lorenz y por tanto existe menor desigualdad.

En los apartados anteriores, ha quedado de manifiesto que existe una serie de propiedades deseables que deben satisfacer los buenos indicadores de desigualdad; sin embargo, no existe consenso en cuáles de ellos deben estar presentes en los estudios que se realicen. No obstante, en aras de orientar el trabajo empírico, es posible argumentar sobre la necesidad de incorporar los siguientes índices.

De los estadísticos que tradicionalmente se utilizan para medir la dispersión de una distribución respecto al valor medio, se sugiere estimar la desigualdad a partir del cálculo del coeficiente de variación al cuadrado y la varianza de los logaritmos. Además, en cualquier estudio sobre la concentración del ingreso, la curva de Lorenz y el coeficiente de Gini resultan imprescindibles para dar una imagen gráfica de la distribución, y para resumir su forma en un indicador sintético de fácil interpretación y gran aceptación, incluso en el público no especializado.

De los índices basados en funciones de bienestar social, el propuesto por Atkinson es el que más relevancia ha adquirido en el trabajo práctico y, de manera específica, se recomienda su cálculo para valores de α=0.5, 1, 1.5, 2 y 2.5.

Por otra parte, también se ha vuelto común la incorporación del Indice de Theil para valores de α =0 y 1. Este indicador tiene la ventaja que se puede descomponer en forma aditiva, y lo por tanto informa sobre la desigualdad al interior de los grupos formados, así como la que existe entre ellos. Además, para el caso en que α=1, el índice de Theil coincide con el coeficiente de Atkinson y si α=2 se obtiene el coeficiente de variación al cuadrado, lo cual permite verificar la consistencia entre los distintos índices comparados.

Además de los indicadores que han sido analizados, algunos de los trabajos académicos, así como los elaborados por organismos nacionales e internacionales, acostumbran incorporar algunos de los siguientes indicadores como medidas de desigualdad: la mediana del ingreso; el porcentaje de ingreso retenido por el decil más bajo de la distribución; el porcentaje de ingreso que se apropia el último decil; el cociente entre los indicadores anteriores; la diferencia absoluta entre los porcentajes de ingreso que concentran el decil más bajo y el más alto; el porcentaje de ingreso retenido por el 40% de los hogares de la parte baja o alta, etc.

A pesar de que estas y otras mediciones pueden ser de utilidad, es importante mencionar que tienen el defecto de no utilizar toda la información asociada a la distribución, y por tanto generan un reporte parcial de lo que ocurre en la población de referencia. Además, en ningún caso aportan más información que los índices que han sido analizados en este trabajo y que utilizan toda la información disponible sobre la población de referencia.

En Cowell (1995) se presenta una taxonomía de las medidas de desigualdad, con relación a tres propiedades que el autor considera fundamentales y que pueden ayudar a decidir sobre la pertinencia de incluir un determinado índice.

CUADRO 9

Indicador de Desigualdad	Principio de Transferencias	Descomposición aditiva	Independencia de escala, ingreso y tamaño de la Población
Varianza	Fuerte	Si	No, aumenta con el ingreso
Coef. de Variación	Débil	Si	Si
Gini	Débil	No*	Si
Atkinson	Débil	Si	Si
Dalton	Débil	Si	No
Theil	Fuerte	Si	Si
Entropía Generalizada	Fuerte	Si	Si

Sólo se satisface en el caso de que el ordenamiento de los grupos de ingresos coincida con el que se obtiene a partir del ingreso total.

Fuente: Tomado de Cowell (1995).

BIBLIOGRAFÍA

Atkinson, (1990)

Cowell, (1995)

Dalton, H. (1920)

Dasgupta, Sen y Starret, (1973)

Efron, B. y Tibshirani, R. (1980). "Bootstrap Methods of Standard Errors, Confidence Intervals and Other Statistical Accuracy", *Statistical Science*, vol. 1, Nº 1, pp. 54-77.

Foster y Sen, (1997)

Kakwani, N. (1980)

Kakwani, N. (1990). "Large Sample Distribution of Several Inequality Measures with Applications to Cote d'Ivore", LSMS *Working Paper* Nº 61, World Bank.

Kakwani, N. (1994)

Medina, F. (1998), "Los Errores de Muestreo en Ecnuestas Complejas", pp. 315-348, en *Memoria del primer Taller Regional del MECOVI*, Aguascalientes, México. CEPAL, 1998.

Pigou, (1912)

Ralphs, (….)

Rothschild y Stiglitz, (1973)

Sen, A. (1976,1992 y 1996)

Shorrocks, (1980 y 1984)