Concentración:
curva de Lorenz e índice de Gini

Estadística Descriptiva

Grau d'Estadística

Manuela Alcañiz Zanón

Ana María Pérez Marín

Jaume Marín Garriga

Dept. Econometria, Estadística i Economia Aplicada

Barcelona, abril de 2018

ÍNDICE

1. Introducción 3
2. Noción de concentración 3
3. Representación gráfica del grado de concentración: curva de Lorenz 6
4. Índice de concentración de Gini 9
5. Ejemplo...12

1. Introducción

La Estadística Descriptiva es la rama de la estadística que recoge, analiza y caracteriza un conjunto de datos, con objeto de describir sus características mediante medidas de resumen, tablas o gráficos.

El primer paso de la descripción de una base de datos consistirá frecuentemente en la elaboración de un gráfico adecuado y, en ocasiones, de una tabla de frecuencias. A continuación, interesará habitualmente estudiar su tendencia central, es decir, encontrar algunas medidas que expresen dónde se halla el "centro" de los datos. También será interesante conocer su dispersión, y la forma de la distribución, que vendrá dada por su grado de asimetría y de apuntamiento.

Si el estudio se centra en datos de determinado tipo de variables (económicas, principalmente) puede resultar de interés el estudio de su concentración. En una distribución estadística de rentas, por ejemplo, desde el punto de vista de la equidad económica, ni la media, ni incluso la varianza son significativas; lo que verdaderamente interesa es la mayor o menor igualdad en su reparto entre los componentes de la población, es decir, el estudio de su equidistribución. Para este fin están concebidos los estudios sobre concentración.

2. Noción de concentración

El término concentración fue introducido en el vocabulario estadístico por el italiano Corrado Gini[1], a propósito de la distribución de salarios y de rentas.

En general, se denomina concentración a la mayor o menor equidad en el reparto de la suma total de la variable considerada. Las variables estadísticas cuya concentración interese estudiar, deberán ser variables positivas y, habitualmente, recogerán magnitudes económicas (rentas, salarios, productividades, etc.).

[1] Gini, C. (1912) Variabilità e Mutuabilità. Contributo allo Studio delle Distribuzioni e delle Relazioni Statistiche. C. Cuppini, Bologna.

Aún cuando "dispersión" y "concentración" tienen significados opuestos en el lenguaje común, el significado estadístico de ambos términos no coincide con el que corrientemente se les otorga.

Desde el punto de vista estadístico, la "dispersión" hace referencia a la variabilidad de los datos, a las diferencias que existen entre ellos y, por tanto, a la mayor o menor representatividad de los promedios.

Las medidas de concentración, sin embargo, tratan de poner en relieve el mayor o menor grado de igualdad en el reparto de la suma total de los valores de una variable. Son, por tanto, indicadores del <u>grado de equidistribución</u> de la variable.

Supongamos que tenemos *n* cantidades que miden los valores de una variable determinada para *n* casos. Para fijar las ideas, supongamos que se trate de la <u>renta</u> de *n* individuos.

Ordenamos dichas cantidades $x_1, x_2, \dots, x_n$, en orden creciente, de modo que cada una de ellas sea menor o igual que la sucesiva, es decir,

$$x_i \leq x_{1+i}, \qquad i = 1, \dots, n-1$$

Nos interesa estudiar hasta qué punto la suma total de rentas $\sum_{i=1}^{n} x_i$ está equitativamente repartida.

Sin duda, las infinitas posiciones que pueden presentarse estarán entre las dos <u>situaciones extremas</u> siguientes:

a) <u>Concentración máxima</u>: de los *n* rentistas, solo uno percibe el total de la renta, en tanto que los demás no perciben nada:

$$x_1 = x_2 = \dots = x_{n-1} = 0, \qquad x_n \neq 0$$

b) <u>Concentración mínima</u> o <u>equidistribución</u>: todos los rentistas perciben la misma cantidad:

$$x_1 = x_2 = \dots = x_n$$

Nos interesa encontrar algunas medidas que permitan valorar cuál es el grado de desigualdad en el reparto de la renta. Para ello, consideremos la siguiente sucesión de rentas acumuladas:

$$S_1 = x_1$$

$$S_2 = x_1 + x_2$$

$$S_3 = x_1 + x_2 + x_3$$

$$\ldots$$

$$S_n = x_1 + x_2 + \ldots + x_n$$

Así, para $i = 1,2,\ldots,n$, S_i es la renta total percibida por los *i* rentistas que menos renta perciben, y S_n es la renta total, cuyo reparto nos interesa estudiar.

Sea q_i el cociente entre S_i y S_n, y p_i el cociente entre *i* y *n*:

$$q_i = S_i / S_n$$

$$p_i = i / n$$

Ello significa que:

q_i = proporción que representa la suma de las *i* rentas inferiores sobre el volumen total de las *n* rentas consideradas = proporción de la renta total que perciben conjuntamente los *i* rentistas con menos renta.

p_i = proporción que representa el número de los *i* rentistas más "pobres" sobre el número total de los rentistas.

Diremos que la concentración de la variable (la renta, en este caso) es tanto más elevada cuanto mayor sea la desigualdad

$$p_i \geq q_i$$

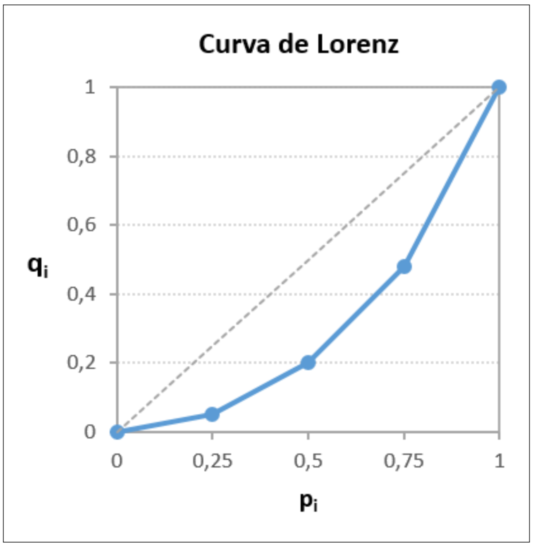
Por ejemplo, si el 80% de los rentistas más "pobres" perciben el 20% de las rentas (y, por tanto, el 20% de los más "ricos" perciben el 80%), el reparto de la renta tendrá más desigualdad que si el 40% de los rentistas más "pobres" perciben el 20% de las rentas.

Es evidente que la desigualdad $p_i \geq q_i$ es cierta en *n-1* casos, o sea, para los *n-1* valores que puede alcanzar *i* desde 1 hasta *n-1*. Sin embargo, cuando *i* alcance el valor *n*, sucede necesariamente que $p_i = q_i$, ya que el total de los rentistas percibe el total de la renta.

Así, para *i* = 1,2,…,*n-1*, siempre se cumplirá que $p_i \geq q_i$, lo que es también fácil de comprender, porque basta imaginar que la sucesión es creciente por hipótesis y, por lo tanto, una fracción determinada de rentistas más pequeños posee una parte del total de la renta que es menos que proporcional al número de rentistas que la poseen. Por ejemplo, la primera mitad de rentistas, formada por los más pequeños, no puede poseer la mitad del volumen total de la renta, sino menos de esa mitad. Existiría proporcionalidad si las cantidades de la sucesión fueran todas iguales.

3. Representación gráfica del grado de concentración: curva de Lorenz

En un diagrama de coordenadas cartesianas representamos sobre el eje de abscisas los valores de p_i, y sobre el eje de ordenadas los valores de q_i. Al unir entre sí los puntos (p_i, q_i), obtendremos una poligonal llamada <u>curva de Lorenz</u>[2]:



Según esta representación, el 25% de los rentistas más "pobres" percibiría el 5% de las rentas; el 50% percibiría el 20%; el 75% percibiría algo menos del 50%; y el 100% percibiría el 100%.

Veamos a continuación algunas <u>propiedades</u> de esta curva de concentración de Lorenz:

[2] Lorenz, M.O. (1905) Methods for Measuring Concentration of Wealth. *Journal of the American Statistical Association*, vol. 9, núm. 70, p. 209-219.

L.1. Es creciente, ya que al considerar proporciones obtenidas de totales acumulados,

$$q_i \leq q_{i+1}, \qquad i = 1,\ldots, n\text{-}1$$

L.2. Se sitúa necesariamente por debajo de la diagonal del cuadrado, ya que las rentas están ordenadas de menor a mayor:

$$x_i \leq x_{i+1}, \qquad i = 1,\ldots, n\text{-}1$$

de modo que es imposible que la proporción p_i de los primeros rentistas supere esa misma proporción en cuanto a volumen de renta acumulada q_i, como ya hemos observado anteriormente.

L.3. Comienza en el origen de coordenadas (0,0) y termina en el punto (1,1): el 0% de los rentistas recibe el 0% de la renta total, y el 100% de ellos percibe el 100% de la renta.

L.4. Es convexa hacia el eje de las abscisas: una fracción determinada de rentistas más pequeños poseerá una fracción de la renta total menor que la fracción poseída por la misma proporción de rentistas mayores, o sea, que una fracción determinada de la renta total es menos que proporcional respecto a la fracción de los rentistas más pequeños que la poseen:

$$q_i - q_{i-1} \leq q_{i+1} - q_i, \qquad i = 2,3,\ldots,n\text{-}1$$

Veamos la demostración:

$$q_i - q_{i-1} = \frac{S_i - S_{i-1}}{S_n} = \frac{x_i}{S_n} \leq \frac{x_{i+1}}{S_n} = \frac{S_{i+1} - S_i}{S_n} = q_{i+1} - q_i$$

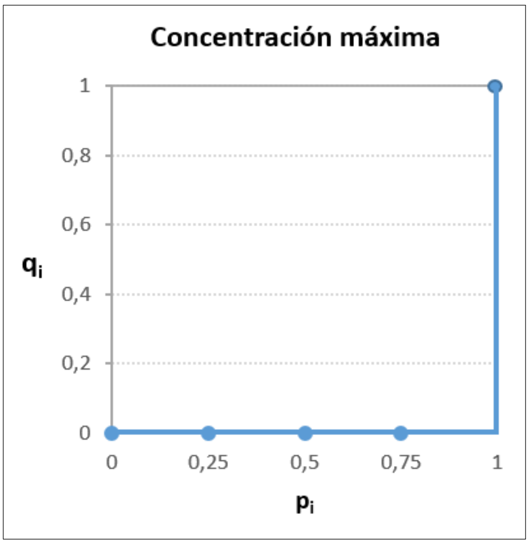
En algunas ocasiones se prefiere cambiar la representación de la curva de Lorenz, poniendo q_i en el eje de abscisas y p_i en el de ordenadas, para mantener la homogeneidad con el criterio de que las frecuencias se sitúan en el eje de ordenadas. En este caso, y en un ejemplo como el nuestro, la curva de Lorenz estaría situada por encima de la diagonal del cuadrado. Sería además creciente y cóncava.

Es interesante ver qué forma adopta la curva de Lorenz en las dos situaciones extremas descritas anteriormente: concentración máxima y equidistribución o reparto igualitario de la variable que se esté considerando (la renta, en este caso).

a) Concentración máxima: un único rentista percibe el total de la renta. De este modo, se verificará forzosamente que:

$$q_1 = q_2 = \dots = q_{n-1} = 0, \qquad q_n = 1$$

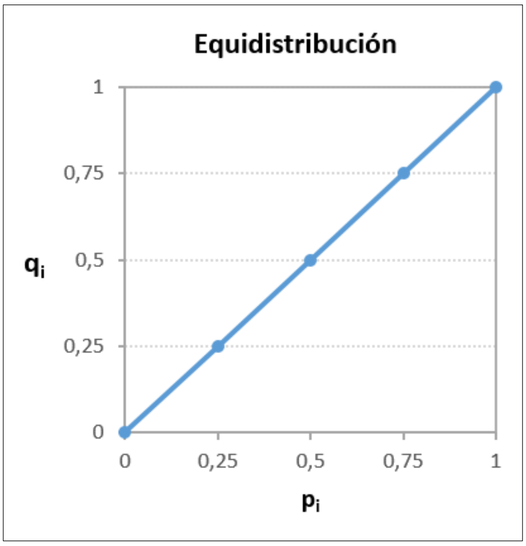
Y así, en el límite ($n \to \infty$), la curva de Lorenz estará formada por los segmentos comprendidos entre (0,0) y (1,0), y entre (1,0) y (1,1):



b) Concentración mínima o equidistribución: todos los rentistas perciben la misma renta individual y, por lo tanto, una fracción determinada de rentistas tiene siempre la misma fracción de la renta total. Entonces, se verificará necesariamente la siguiente igualdad:

$$p_i = q_i, \qquad i = 1, \dots, n$$

con lo que la curva de Lorenz se transformará en la diagonal del cuadrado:



Este es, al igual que el anterior, un caso límite, ya que habitualmente las cantidades serán desiguales, es decir, los rentistas poseerán una renta individual distinta.

A la vista de las dos representaciones anteriores, podemos empezar a deducir que cuanto más próxima se halle la curva de Lorenz a la diagonal del cuadrado, tanto más equidistribuida estará la magnitud que se esté considerando, y menos nivel de concentración habrá.

Será interesante, sin embargo, estudiar algunas medidas cuantitativas que nos permitan valorar el grado de concentración de un modo más preciso que la curva de Lorenz. Nos centraremos en la medida más conocida, que es el índice de concentración de Gini, si bien existen otras, como el índice de Theil.

4. Índice de concentración de Gini

Hemos dicho que pueden establecerse *n-1* desigualdades $p_i \geq q_i$ y que de sus valores depende la intensidad de la concentración. Podemos, por lo tanto, obtener un índice de concentración a partir de una fórmula que tenga en cuenta estas *n-1* desigualdades o, de forma equivalente, las *n-1* diferencias $p_i - q_i$. La fórmula más sencilla que satisface estas condiciones es, sin duda, la siguiente:

$$\sum_{i=1}^{n-1}(p_i - q_i)$$

Cuanto mayor sea el valor de esta suma, tanto mayor será la concentración. Para mayor comodidad en las comparaciones, será conveniente obtener un índice que tenga un valor máximo igual a la unidad. Bastará para ello dividir la suma anterior por el valor máximo que pueda alcanzar, que corresponderá al caso:

$$q_1 = q_2 = \ldots = q_{n-1} = 0, \quad q_n = 1$$

es decir, a la situación de <u>concentración máxima</u> (notemos que q_n no interviene en el sumatorio). Este valor máximo será, pues:

$$\sum_{i=1}^{n-1} p_i$$

Y el <u>índice de concentración de Gini</u> quedará definido como:

$$G = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

Este índice, muy utilizado en la práctica, verifica las siguientes propiedades:

G.1. G crece con el aumento de la concentración: al aumentar la desigualdad, las diferencias $p_i - q_i$ se hacen mayores.

G.2. En el caso de concentración máxima, G = 1: es así por construcción, ya que entonces:

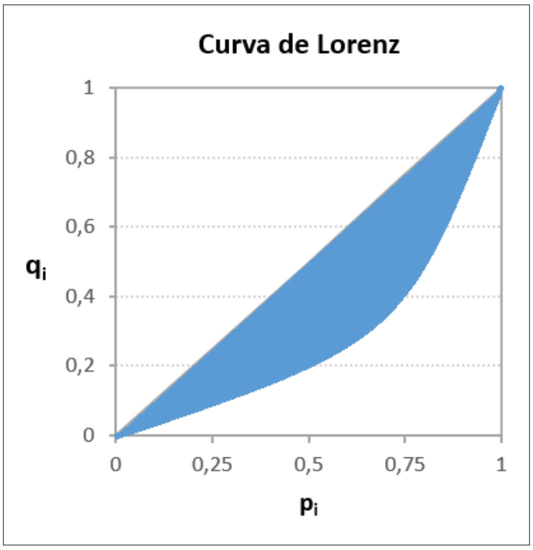
$$q_1 = q_2 = \ldots = q_{n-1} = 0$$

G.3. Cuando la concentración es nula, es decir, cuando hay equidistribución, G =0: es evidente, ya que entonces:

$$p_i = q_i, \quad i = 1,\ldots,n$$

G.4. Cualquier transferencia de un individuo "rico" a otro más "pobre", *ceteris paribus,* reduce G. Del mismo modo, toda transferencia de un individuo "pobre" a otro más "rico", hace que G aumente (condición de Pigou-Dalton).

Por otra parte, observemos el área pintada en la siguiente figura:

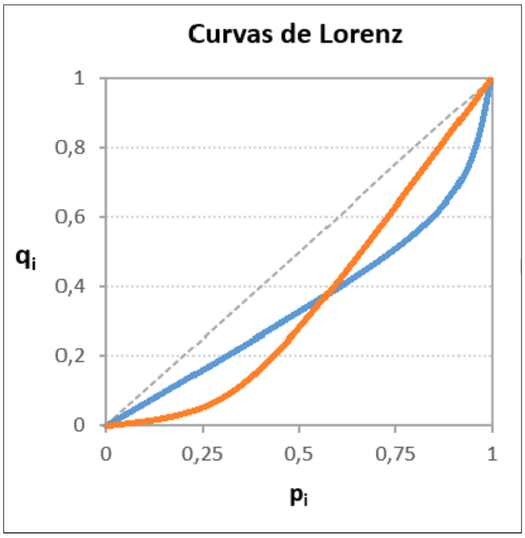


Se puede demostrar fácilmente de forma empírica que el índice de Gini es aproximadamente igual al área comprendida entre la línea de equidistribución (diagonal del cuadrado) y la curva de concentración (área rayada), dividida por el área del triángulo inferior a la diagonal.

Dado que el área del triángulo es igual a 0,5, el índice de Gini es aproximadamente igual al doble del área pintada en la figura anterior. Esta área puede medirse de forma bastante precisa mediante el método numérico de los trapecios[3].

Por último, cabe señalar que, si bien el índice de Gini tiene la ventaja de resumir en una sola cifra las complejas informaciones expresadas por la curva de Lorenz, y por lo mismo permite comparar más fácilmente que la curva la concentración de dos distribuciones, esta ventaja tiene su contrapartida: dos distribuciones de aspectos muy diferentes pueden, en efecto, tener dos índices de concentración de Gini del mismo valor. Así, las distribuciones representadas por las siguientes curvas de Lorenz tienen el mismo grado de concentración global, aunque la estructura del reparto de la variable no es la misma. En el caso de la curva naranja, la situación es más desfavorable para los rentistas más “pobres”:

[3] Iriarte, R. (2007) Métodos numéricos. Editorial Trillas.



5. Ejemplo

Dos empresas A y B tienen 50 trabajadores cada una. Las dietas de desplazamiento dependen del trabajador. Concretamente, son las siguientes:

Dieta (euros)	Trabajadores empresa A	Trabajadores empresa B
80	10	5
60	5	15
20	5	15
15	5	10
10	25	5

Con objeto de ver en qué empresa hay más concentración en las dietas, representaremos las curvas de Lorenz y calcularemos los índices de Gini para cada una.

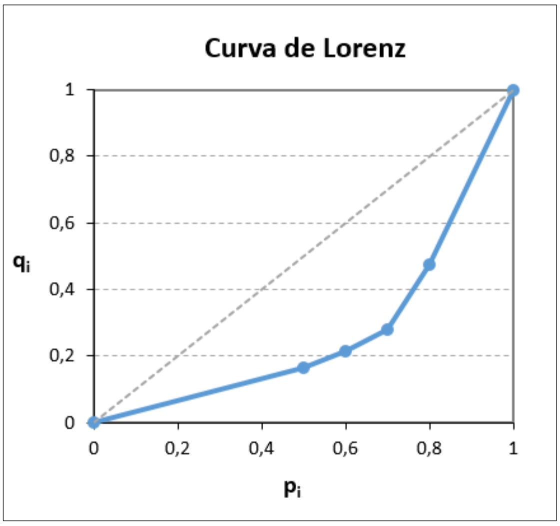
Conviene empezar construyendo la siguiente tabla, recordando que los valores x_i siempre deben estar ordenados de menor a mayor. En caso contrario, los resultados serán erróneos.

Empresa A:

x_i	n_i	N_i	$n_i x_i$	S_i	p_i	q_i	p_i-q_i
10	25	25	250	250	50	16,39	33,61
15	5	30	75	325	60	21,31	38,69
20	5	35	100	425	70	27,87	42,13
60	5	40	300	725	80	47,54	32,46
80	10	50	800	1525	100	100,00	0,00

n_i=frecuencia absoluta, N_i=frecuencia absoluta acumulada

La tabla permite dibujar la curva de Lorenz, y calcular el índice de Gini:



$$G = \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i} = \frac{33{,}61 + 38{,}69 + 42{,}13 + 32{,}46}{50 + 60 + 70 + 80} = 0{,}56$$

Se concluye que la empresa A tiene un grado de concentración considerable en la repartición de las dietas de desplazamiento. Observemos que un valor de 0,56 está indicando una concentración elevada. Es muy infrecuente en la práctica que G alcance valores cercanos a 1.

<u>Ejercicio</u>: repite los cálculos con ayuda de Excel para la empresa B. Comprueba que el índice de Gini vale 0,37. Dibuja también la curva de Lorenz de la empresa B y compárala con la de la empresa A. ¿Qué se concluye?