MEDIDAS DE FORMA Y CONCENTRACIÓN

4.1.- Asimetría: coeficientes de asimetría de Fisher y Pearson. Otros Coeficientes de asimetría.

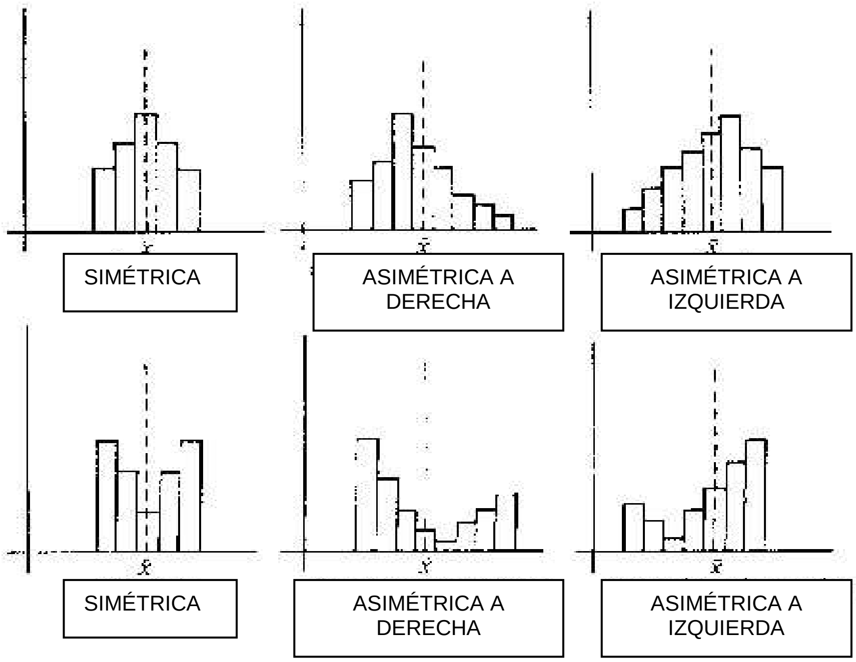
Medidas de forma:

- Asimetría
- Curtosis o apuntamiento.

Hasta ahora, hemos estado analizando y estudiando la dispersión de una distribución, pero parece evidente que necesitamos conocer más sobre el comportamiento de una distribución. En esta parte, analizaremos las medidas de forma, en el sentido de histograma o representación de datos, es decir, que información nos aporta según la forma que tengan la disposición de datos.

Las medidas de forma de una distribución se pueden clasificar en dos grandes grupos o bloques: *medidas de asimetría y medidas de curtosis*.

Cuando al trazar una vertical, en el diagrama de barras o histograma, de una variable, según sea esta discreta o continua, por el valor de la media, esta vertical, se transforma en eje de simetría, decimos que la distribución es simétrica. Diremos pues, que es simétrica, cuando a ambos lados de la media aritmética haya el mismo nº de valores de la variable, equidistantes de dicha media dos a dos, y tales que cada par de valores equidistantes tiene la misma frecuencia absoluta. En caso contrario, dicha distribución será asimétrica o diremos que presenta asimetría.

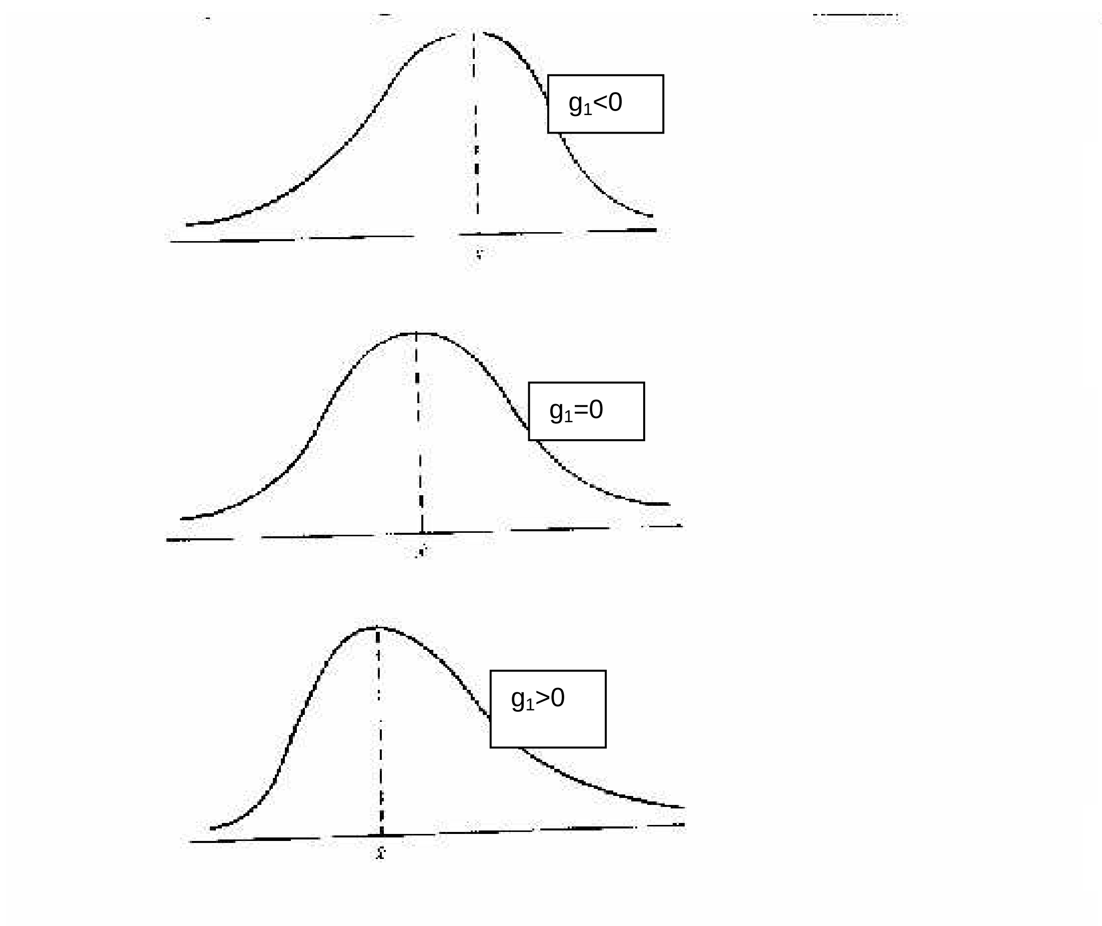


Para calcular la asimetría, una posibilidad, es utilizar el llamado ***coeficiente de FISHER*** que representaremos como $\mathbf{g_1}$ y responderá a la siguiente expresión matemática:

$$\boxed{g_1 = \frac{\sum (x_i - \bar{x})^3 n_i}{ns^3}}$$

Según sea el valor de g_1, diremos que la distribución es asimétrica a derechas o positiva, a izquierdas o negativa, o simétrica, o sea:

Si $\mathbf{g_1 > 0}$ ➔ *la distribución será asimétrica positiva o a derechas (desplazada hacia la derecha).*
Si $\mathbf{g_1 < 0}$ ➔ *la distribución será asimétrica negativa o a izquierdas (desplazada hacia la izquierda).*
Si $\mathbf{g_1 = 0}$ ➔ *la distribución será simétrica*.



Otra posibilidad de calcular la asimetría, es por medio del ***coeficiente de PEARSON (Ap)***, el cual responde a la siguiente expresión.

$$\boxed{A_p = \frac{\overline{X} - Mo}{S}}$$

Aunque en la práctica este coeficiente sería más fácil de calcular que el anterior, casi no lo utilizaremos ya que solo es cierto cuando la distribución tiene las siguientes condiciones:

Unimodal
Campaniforme
Moderada o ligeramente asimetrica.

Si Ap > 0 ➔ *la distribución será asimétrica positiva o a derechas (desplazada hacia la derecha).*
Si Ap < 0 ➔ *la distribución será asimétrica negativa o a izquierdas (desplazada hacia la izquierda).*
Si Ap = 0 ➔ *la distribución será simétrica.*

NOTA: Otro coeficiente es el **coeficiente de asimetría de Bowley**, menos utilizado. El cual esta basado en la posición de los cuartiles y la mediana, para lo cual los relacionaremos de acuerdo con la siguiente expresión:

$$\boxed{Ab = \frac{C_3 + C_1 - 2Me}{C_3 + C_1}}$$

4.2.- La ley normal.

Se hace necesario, para la teoría siguiente, conocer la ***DISTRIBUCIÓN NORMAL***, ya que tiene gran importancia al querer estudiar el apuntamiento o curtosis. Se dice que una distribución tiene un apuntamiento u otro, siempre en función de esta distribución normal.

La distribución llamada normal, corresponde a fenómenos muy corrientes en la naturaleza y cuya representación gráfica es una campana de Gauss. Esta campana responde a una función matemática, que es la función de densidad de la distribución:

$$f(x)=\frac{1}{s\sqrt{2\pi}}e^{-\frac{(x-\overline{X})^2}{2S^2}}$$

Se producen unos punto de inflexión X +S y X-S y el eje OX es una asíntota horizontal siendo el área comprendida entre la f y el eje de las X igual a 1

4.3.- Curtosis: coeficiente de Fisher.

Para calcularlo utilizaremos la expresión

$$\boxed{g_2=\frac{\sum(x_i-\overline{X})^4 n_i}{ns^4}-3}$$

Si $\mathbf{g_2 > 0}$ *la distribución será leptocúrtica o apuntada*
Si $\mathbf{g_2 = 0}$ *la distribución será mesocúrtica o normal*
Si $\mathbf{g_2 < 0}$ *la distribución será platicúrtica o menos apuntada que lo normal*.

4.4.- Medidas de concentración: Indice de Gini y Curva de Lorenz.

Las medidas de concentración tratan de poner de relieve el mayor o menor grado de igualdad en el reparto del total de los valores de la variable, son por tanto indicadores del grado de distribución de la variable.

Para este fin, están concebidos los estudios sobre concentración.

Denominamos concentración a la *mayor o menor equidad en el reparto de la suma total de los valores de la variable considerada (renta, salarios, etc.)*.

Las infinitas posibilidades que pueden adoptar los valores, se encuentran entre los dos extremos:

1.- **Concentración máxima,** cuando uno solo percibe el total y los demás nada, en este caso, nos encontraremos ante un reparto no equitativo:

$x_1 = x_2 = x_3 = \ldots\ldots\ldots\ldots = x_{n-1} = 0$ y x_n.

2.- **Concentración mínima,** cuando el conjunto total de valores de la variable esta repartido por igual, en este caso diremos que estamos ante un reparto equitativo

$x_1 = x_2 = x_3 = \ldots\ldots\ldots\ldots = x_{n-1} = x_n$

De las diferentes medidas de concentración que existen nos vamos a centrar en dos:

Indice de Gini, Coeficiente, por tanto será un valor numérico.

Curva de Lorenz, gráfico, por tanto será una representación en ejes coordenados.

Sea una distribución de rentas (x_i, n_i) de la que formaremos una tabla con las siguientes columnas:

1.- Los productos $\mathbf{x_i\, n_i}$, que nos indicarán la renta total percibida por los $\mathbf{n_i}$ rentistas de renta individual $\mathbf{x_i}$.

2.- Las frecuencias absolutas acumuladas $\mathbf{N_i}$.

3.- Los totales acumulados $\mathbf{u_i}$ que se calculan de la siguiente forma:

$$u_1 = x_1\, n_1$$
$$u_2 = x_1\, n_1 + x_2\, n_2$$
$$u_3 = x_1\, n_1 + x_2\, n_2 + x_3\, n_3$$
$$u_4 = x_1\, n_1 + x_2\, n_2 + x_3\, n_3 + x_4\, n_4$$
$$u_n = x_1\, n_1 + x_2\, n_2 + x_3\, n_3 + x_4\, n_4 + \ldots\ldots\ldots\ldots + x_n\, n_n$$

Por tanto podemos decir que $u_n = \sum_{i=1}^{n} x_i n_i$

4.- La columna total de frecuencias acumuladas relativas, que expresaremos en tanto por ciento y que representaremos como $\mathbf{p_i}$ y que vendrá dada por la siguiente notación

$$\boxed{p_i = \frac{N_i}{n} 100}$$

5.- La renta total de todos los rentistas que será $\mathbf{u_n}$ y que dada en tanto por ciento, la cual representaremos como $\mathbf{q_i}$ y que responderá a la siguiente notación:

$$\boxed{q_i = \frac{u_i}{u_n} 100}$$

Por tanto ya podemos confeccionar la tabla que será la siguiente:

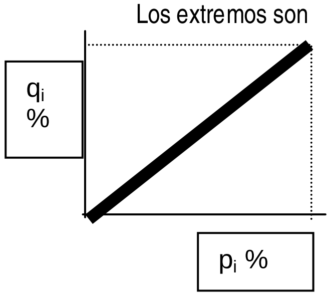
x_i	n_i	$x_i\, n_i$	N_i	u_i	$p_i = \frac{N_i}{n} 100$	$q_i = \frac{u_i}{u_n} 100$	$p_i - q_i$
x_1	n_1	$x_1\, n_1$	N_1	u_1	p_1	q_1	$p_1 - q_1$
x_2	n_2	$x_2\, n_2$	N_2	u_2	p_2	q_2	$p_2 - q_2$
...	...	...	...	...	...	...	...
x_n	n_n	$x_n\, n_n$	N_n	u_n	p_n	q_n	$p_n - q_n$

Como podemos ver la última columna es la diferencia entre las dos penúltimas, esta diferencia seria 0 para la concentración mínima ya que $p_i = q_i$ y por tanto su diferencia seria cero.

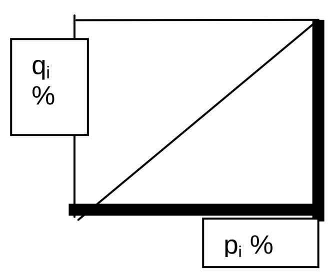
Si esto lo representamos gráficamente obtendremos la **curva de concentración o curva de Lorenz** .La manera de representarlo será, en el eje de las X, los valores p_i en % y en el de las Y los valores de q_i en %. Al ser un %, el gráfico siempre será un cuadrado, y la gráfica será una curva que se unirá al cuadrado, por los valores (0,0), y (100,100), y quedará siempre por debajo de la diagonal.

La manera de interpretarla será: cuanto más cerca se sitúe esta curva de la diagonal, menor concentración habrá, o más homogeneidad en la distribución. Cuanto más se acerque a los ejes, por la parte inferior del cuadrado, mayor concentración.

Los extremos son



Distribución de concentración mínima



Distribución de concentración máxima

Analíticamente calcularemos el **índice de Gini** el cual responde a la siguiente ecuación

$$I_G = \frac{\sum_{i=1}^{k-1}(p_i - q_i)}{\sum_{i=1}^{k-1} p_i}$$

Este índice tomara los valores de **I_G = 0 cuando $p_i = q_i$ concentración mínima** y de **I_g = 1 cuando q_i = 0**
Esto lo veremos mejor con un ejemplo

		Frecuencia						
L_{i-1} - L_i	**marca x_i**	**n_i**	**N_i**	**$x_i n_i$**	**Σu_n**	**qi = (u_i/u_n) 100**	**pi = (Ni/n) 100**	**p_i - q_i**
0 - 50	25	23	23	575	575	1,48	8,85	7,37
50 - 100	75	72	95	5400	5975	15,38	36,54	21,16
100 - 150	125	62	157	7750	13725	35,33	60,38	25,06
150 - 200	175	48	205	8400	22125	56,95	78,85	21,90
200 - 250	225	19	224	4275	26400	67,95	86,15	18,20
250 - 300	275	8	232	2200	28600	73,62	89,23	15,61
300 - 350	325	14	246	4550	33150	85,33	94,62	9,29
350 - 400	375	7	253	2625	35775	92,08	97,31	5,22
400 - 450	425	5	258	2125	37900	97,55	99,23	1,68
450 - 500	475	2	260	950	38850	100,00	100,00	0,00
		260		38850			651,15	125,48

Se pide Indice de concentración y Curva de Lorenz correspondiente

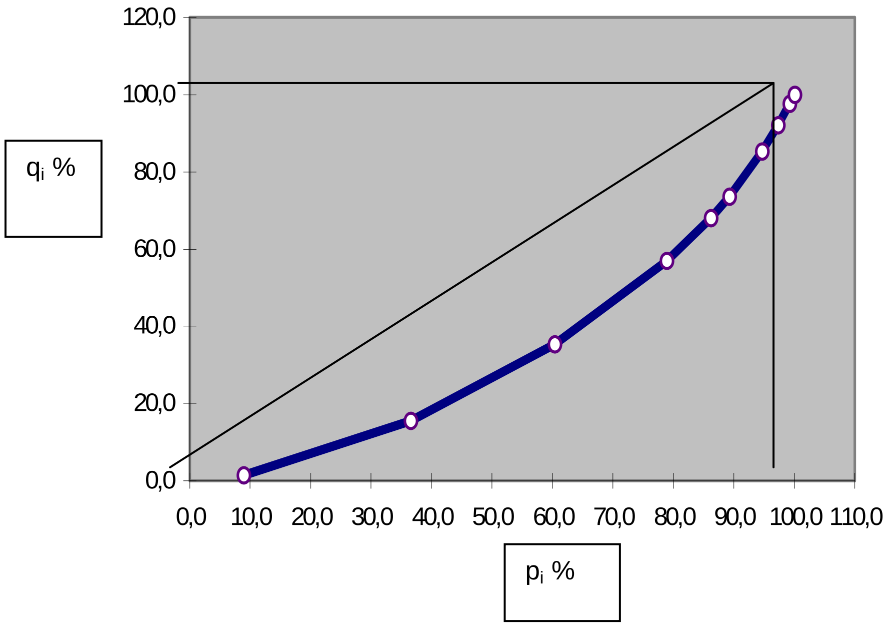
a) **Indice de concentración de GINI**

$$I_G = \frac{\sum_{i=1}^{k-1}(p_i - q_i)}{\sum_{i=1}^{k-1} p_i} = \frac{125{,}48}{651{,}15} = 0{,}193 \text{ , Observamos}$$

que hay poca concentración por encontrarse cerca del 0.

b) Curva de Lorenz

La curva la obtenemos cerca de la diagonal, que indica que hay poca concentración:


120,0
100,0
80,0
60,0
40,0
20,0
0,0
0,0 10,0 20,0 30,0 40,0 50,0 60,0 70,0 80,0 90,0 100,0 110,0
qi %
pi %