Capítulo II	CARACTERISTICAS DE UNA DISTRIBUCIÓN DE FRECUENCIAS

2.1. Introducción

La fase previa de cualquier estudio estadístico se basa en la recogida y ordenación de datos; esto se realiza con la ayuda de los resúmenes numéricos y gráficos visto en los temas anteriores.

2.2. Medidas de posición

Son aquellas medidas que nos ayudan a saber *donde* están los datos pero sin indicar como se distribuyen.

2.2.1. Medidas de posición central

a) Media aritmética ($\overline{X}$)

La media aritmética o simplemente media, que denotaremos por $\overline{X}$, es el número obtenido al dividir la suma de todos los valores de la variable entre el numero total de observaciones, y se define por la siguiente expresión:

$$\boxed{\overline{x} = \frac{\sum_{i=1}^{n} x_i n_i}{N}}$$

Ejemplo:

Si tenemos la siguiente distribución, se pide hallar la media aritmética, de los siguientes datos expresados en kg.

xi	ni	xi ni
54	2	108
59	3	177
63	4	252
64	1	64
	N=10	601

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i n_i}{N} = \frac{601}{10} = \underline{\underline{60{,}1}} \quad \text{kg}$$

Si los datos están agrupados en intervalos, la expresión de la media aritmética, es la misma, pero utilizando la marca de clase (Xi).

Ejemplo:

$(L_{i-1}, L_i]$	x_i	n_i	$x_i n_i$

[30 , 40]	35	3	105
(40 , 50]	45	2	90
(50 , 60]	55	5	275
		10	470

$$\overline{X} = \frac{\sum_{i=1}^{n} x_i n_i}{N} = \frac{470}{10} = \underline{\underline{47}}$$

Propiedades:
1ª) Si sometemos a una variable estadística X, a un cambio de origen y escala Y = a + b X, la media aritmética de dicha variable X, varía en la misma proporción.

$$Y = a + bX \longrightarrow \overline{Y} = a + b\overline{X}$$

2ª) La suma de las desviaciones de los valores o datos de una variable X, respecto a su media aritmética es cero.

$$\boxed{\sum_{i=1}^{n} (x_i - \overline{x}) n_i = 0}$$

Ventajas e inconvenientes:
- La media aritmética viene expresada en las mismas unidades que la variable.
- En su cálculo intervienen todos los valores de la distribución.
- Es el centro de gravedad de toda la distribución, representando a todos los valores observados.
- Es única.
- Su principal inconveniente es que se ve afectada por los valores extremadamente grandes o pequeños de la distribución.

- Media aritmética ponderada

Es una media aritmética que se emplea en distribuciones de tipo unitario, en las que se introducen unos coeficientes de ponderación, denominados ω_i, que son valores positivos, que representan el número de veces que un valor de la variable es más importante que otro.

$$W = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$$

b) <u>Media geométrica</u>

Sea una distribución de frecuencias (x_i, n_i). La media geométrica, que denotaremos por G. se define como la raíz N-ésima del producto de los N valores de la distribución.

G = $\sqrt[N]{x_1^{n1} x_2^{n2} \cdots\cdots x_k^{nk}}$

Si los datos están agrupados en intervalos, la expresión de la media geométrica, es la misma, pero utilizando la marca de clase (Xi).

El empleo más frecuente de la media geométrica es el de promediar variables tales como porcentajes, tasas, números índices. etc., es decir, en los casos en los que se supone que la variable presenta variaciones acumulativas.

Ventajas e inconvenientes:
- En su cálculo intervienen todos los valores de la distribución.
- Los valores extremos tienen menor influencia que en la media aritmética.
- Es única.
- Su cálculo es más complicado que el de la media aritmética.

Además, cuando la variable toma al menos un $x_i = 0$ entonces G se anula, y si la variable toma valores negativos se pueden presentar una gama de casos particulares en los que tampoco queda determinada debido al problema de las raíces de índice par de números negativos.

c) Media armónica

La media armónica, que representaremos por H, se define como sigue:

$$\boxed{H = \frac{N}{\sum_{i=1}^{r} \frac{1}{x_i} n_i}}$$

Obsérvese que la inversa de la media armónica es la media aritmética de los inversos de los valores de la variable. No es aconsejable en distribuciones de variables con valores pequeños. Se suele utilizar para promediar variables tales como productividades, velocidades, tiempos, rendimientos, cambios, etc.

Ventajas e inconvenientes:
- En su cálculo intervienen todos los valores de la distribución.
- Su cálculo no tiene sentido cuando algún valor de la variable toma valor cero.
- Es única.

- Relación entre las medias:

$$\boxed{H \leq G \leq \overline{X}}$$

d) Mediana (Me)

Dada una distribución de frecuencias con los valores ordenados de menor a mayor, llamamos mediana y la representamos por Me, al valor de la variable, que deja a su izquierda el mismo número de frecuencias que a su derecha.

- Calculo de la mediana:

Variara según el tipo de dato:

a) Variables discretas no agrupadas:

1º) Se calcula $\frac{N}{2}$ y se construye la columna de las Ni (frecuencias acumuladas)

2º) Se observa cual es la primera Ni que supera o iguala a $\frac{N}{2}$, distinguiéndose dos casos:

- Si existe un valor de Xi tal que $N_{i-1} \prec \frac{N}{2} \prec N_i$, entonces se toma como $\boxed{Me = x_i}$
- Si existe un valor i tal que $N_i = \frac{N}{2}$, entonces la $\boxed{Me = \frac{x_i + x_{i+1}}{2}}$

Ejemplo: Sea la distribución

xi	ni	Ni
1	3	3
2	4	7
5	9	16
7	10	26
10	7	33
13	2	35
	n = 35	

lugar que ocupa $\frac{N}{2} = \frac{35}{2} = 17{,}5$

como se produce que $N_{i-1} < \frac{N}{2} < N_i \Rightarrow 16 < 17{,}7 < 26 \Rightarrow Me = x_i$,por lo tanto Me = 7

El otro caso lo podemos ver en la siguiente distribución:

xi	ni	Ni
1	3	3
2	4	7

5	9	16
7	10	26
10	6	32
	n= 32	

Lugar que ocupa = 32/2 = 16 ==> $Me = \frac{x_1 + x_{i+1}}{2} = \frac{5+7}{2} = 6$

Notar que en este caso se podría haber producido que hubiera una frecuencia absoluta acumulada superior a 16. En este caso se calcularía como en el ejemplo anterior.

b) Variables agrupadas por intervalos

En este caso hay que detectar en que intervalo está el valor mediano. Dicho intervalo se denomina " intervalo mediano ".

Cada intervalo I_i vendrá expresado según la notación $I_i = (L_{i-1}, L_i]$; observando la columna de las frecuencias acumuladas, buscaremos el primer intervalo cuya Ni sea mayor o igual que $\frac{N}{2}$, que será el intervalo modal; una vez identificado dicho intervalo, procederemos al cálculo del valor mediano, debiendo diferenciar dos casos:

1º) Si existe Ii tal que $N_{i-1} \prec \frac{N}{2} \prec N_i$, entonces el intervalo mediano es el $(L_{i-1}, L_i]$ y la mediana es:

$$\boxed{M_e = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} c_i}$$

2º) Análogamente si existe un Ii tal que $N_i = \frac{N}{2}$, la mediana es $\boxed{Me = L_i}$

Ejemplo:

(Li-1, Li]	ni	Ni
[20 , 25]	100	100
(25 , 30]	150	250
(30 , 35]	200	450
(35 , 40]	180	630
(40 , 45]	41	671
	N = 671	

671/2 = 335.5 ; Me estará en el intervalo (30 - 35]. Por tanto realizamos el cálculo:

$$Me = L_{i-1} + \frac{\frac{N}{2} - N_{i-1}}{n_i} a_i = 30 + \frac{33{,}5 - 250}{200} * 5 = \underline{\underline{32{,}138}}$$

Ventajas e inconvenientes :

- Es la medida más representativa en el caso de variables que solo admitan la escala ordinal.
- Es fácil de calcular.
- En la mediana solo influyen los valores centrales y es insensible a los valores extremos u "outliers ".
- En su determinación no intervienen todos los valores de la variable.

e) Moda

La moda es el valor de la variable que más veces se repite, y en consecuencia, en una distribución de frecuencias, es el valor de la variable que viene afectada por la máxima frecuencia de la distribución. En distribuciones no agrupadas en intervalos se observa la columna de las frecuencias absolutas, y el valor de la distribuci6n al que corresponde la mayor frecuencia será la moda. A veces aparecen distribuciones de variables con más de una moda (bimodales, trimodales, etc), e incluso una distribución de frecuencias que presente una moda absoluta y una relativa.

En el caso de estar la variable agrupada en intervalos de distinta amplitud, se define el intervalo modal, y se denota por (L_{i-1} , L_i], como aquel que posee mayor densidad de frecuencia (h_i); la densidad de frecuencia se define como : $\boxed{h_i = \frac{n_i}{a_i}}$

Una vez identificado el intervalo modal procederemos al cálculo de la moda, a través de la fórmula:

$$\boxed{Mo = L_{i-1} + \frac{h_{i+1}}{h_{i-1} + h_{i+1}} c_i}$$

En el caso de tener todos los intervalos la misma amplitud, el intervalo modal será el que posea una mayor frecuencia absoluta (n_i) y una vez identificado este, empleando la fórmula:

$$\boxed{Mo = L_{i-1} + \frac{n_{i+1}}{n_{i-1} + n_{i+1}} c_i}$$

Ventajas e inconvenientes:

- Su cálculo es sencillo.
- Es de fácil interpretación.

- Es la única medida de posición central que puede obtenerse en las variables de tipo cualitativo.
- En su determinación no intervienen todos lo valores de la distribución.

2.2.2. **Medidas de posición no central (Cuantiles)**

Los cuantiles son aquellos valores de la variable, que ordenados de menor a mayor, dividen a la distribución en partes, de tal manera que cada una de ellas contiene el mismo número de frecuencias.

Los cuantiles más conocidos son:

a) Cuartiles (Qi)

Son valores de la variable que dividen a la distribución en 4 partes, cada una de las cuales engloba el 25 % de las mismas. Se denotan de la siguiente forma: Q_1 es el primer cuartil que deja a su izquierda el 25 % de los datos; Q_2 es el segundo cuartil que deja a su izquierda el 50% de los datos, y Q_3 es el tercer cuartil que deja a su izquierda el 75% de los datos. (Q_2 = Me)

b) Deciles (Di)

Son los valores de la variable que dividen a la distribución en las partes iguales, cada una de las cuales engloba el 10 % de los datos. En total habrá 9 deciles. ($Q_2 = D_5$ = Me)

c) Centiles o Percentiles (Pi)

Son los valores que dividen a la distribución en 100 partes iguales, cada una de las cuales engloba el 1 % de las observaciones. En total habrá 99 percentiles. ($Q_2 = D_5 = Me = P_{50}$)

- **Cálculo de los cuantiles en distribuciones no agrupadas en intervalos**

- Se calculan a través de la siguiente expresión: $\boxed{\frac{rN}{q}}$, siendo :

r = el orden del cuantil correspondiente

q = el número de intervalos con iguales frecuencias u observaciones (q = 4, 10, ó 100).

N = número total de observaciones

- La anterior expresión nos indica que valor de la variable estudiada es el cuantil que nos piden, que se corresponderá con el primer valor cuya frecuencia acumulada sea mayor o igual a $\frac{rN}{q}$

Ejemplo: DISTRIBUCIONES NO AGRUPADAS: En la siguiente distribución

xi	ni
5	3
10	7
15	5
20	3
25	2
	N = 20

Ni
3
10
15
18
20

Calcular la mediana (Me); el primer y tercer cuartil (C1,C3); el 4º decil (D4) y el 90 percentil (P90)

Mediana (Me)

Lugar que ocupa la mediana ➔ lugar 20/2 = 10

Como es igual a un valor de la frecuencia absoluta acumulada, realizaremos es cálculo: $Me = \frac{x_i + x_{i+1}}{2} = \frac{10+15}{2} = \underline{\underline{12{,}5}}$

Primer cuartil (C1)

Lugar que ocupa en la distribución (¼). 20 = 20/4 = 5 Como Ni-1 < $\frac{rN}{q}$ < Ni , es decir 3 < 5 < 10 esto implicara que C1 = xi = 10

Tercer cuartil (C3)

Lugar que ocupa en la distribución (3/4).20 = 60/4 = 15, que coincide con un valor de la frecuencia absoluta acumulada, por tanto realizaremos el cálculo: $C_3 = \frac{x_i + x_{i+1}}{2} = \frac{15+20}{2} = \underline{\underline{17{,}5}}$

Cuarto decil (D4)

Lugar que ocupa en la distribución (4/10) . 20 = 80/10 = 8. Como Ni-1 < $\frac{rN}{q}$ < Ni ya que 3 < 8 < 10 por tanto D4 =10.

Nonagésimo percentil (P90)

Lugar que ocupa en la distribución (90/100). 20 = 1800/100 = 18. que coincide con un valor de la frecuencia absoluta acumulada, por tanto realizaremos el cálculo: $P_{90} = \frac{x_i + x_{i+1}}{2} = \frac{20+25}{2} = \underline{\underline{22{,}5}}$

- **Cálculo de los cuantiles en distribuciones agrupadas en intervalos**

- Este cálculo se resuelve de manera idéntica al de la mediana.
- El intervalo donde se encuentra el cuantil i-esimo, es el primero que una vez ordenados los datos de menor a mayor, tenga como frecuencia acumulada (Ni) un valor superior o igual a $\frac{rN}{q}$; una vez

identificado el intervalo Ii (Li-1 , Li], calcularemos el cuantil correspondiente, a través de la fórmula:

$$C_{r/q} = L_{i-1} + \frac{\frac{rN}{q} - N_{i-1}}{n_i} c_i$$

r=1,2,...,q-1.

Cuartil: q=4; Decil: q=10; Percentil: q=100

Ejemplo:

DISTRIBUCIONES AGRUPADAS: Hallar el primer cuartil, el cuarto decil y el 90 percentil de la siguiente distribución:

[Li-1 , Li)	ni	Ni
[0 , 100]	90	90
(100 , 200]	140	230
(200 , 300]	150	380
(300 , 800]	120	500
	N = 500	

- Primer cuartil (Q1)
- Lugar ocupa el intervalo del primer cuartil: (1/4). 500 = 500/4 = 125. Por tanto Q1 estará situado en el intervalo (100 – 200].Aplicando la expresión directamente, tendremos: $Q_1 = 100 + \frac{125-90}{140}100 = \underline{\underline{125}}$

- Cuarto decil (D4)
- Lugar que ocupa: (4/10) . 500 = 200 . Por tanto D4 estará situado en el intervalo (100 – 200]. Aplicando la expresión tendremos: $D_4 = 100 + \frac{200-90}{140}100 = \underline{\underline{178{,}57}}$
-
- Nonagésimo percentil (P 90)
- Lugar que ocupa: (90/100) . 500 = 450, por tanto P90 estará situado en el intervalo (300 – 800]. Aplicando la expresión tendremos: $P_{90} = 300 + \frac{450-380}{120}500 = 300 + \frac{70}{120}500 = \underline{\underline{591{,}67}}$

2.3. Momentos potenciales

Los momentos son medidas obtenidas a partir de todos los datos de una variable estadística y sus frecuencias absolutas. Estas medidas caracterizan a las distribuciones de frecuencias de tal forma que si los momentos coinciden en dos distribuciones, diremos que son iguales.

2.3.1. Momentos respecto al origen

Se define el momento de orden h respecto al origen de una variable estadística a la expresión:

$$\boxed{a_h = \frac{\sum_{i=1}^{n} x_i^{\,h} n_i}{N}}$$

Particularidades:

Si h = 1, a1 es igual a la media aritmética.
Si h = 0, a0 es igual a uno (a0 = 1)

2.3.2. Momentos centrales o momentos con respecto a la media aritmética

$$\boxed{m_h = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^h n_i}{N}}$$

Particularidades:

- Si h = 1, entonces m1 = 0
- Si h = 2, entonces m2 = S^2

2.4. Medidas de dispersión

Las medidas de dispersión tratan de medir el grado de dispersión que tiene una variable estadística en torno a una medida de posición o tendencia central, indicándonos lo representativa que es la medida de posición. A mayor dispersión menor representatividad de la medida de posición y viceversa.

2.4.1 **Medidas de dispersión absoluta**

a) Recorrido (Re)

Se define como la diferencia entre el máximo y el mínimo valor de la variable:

$$R = \text{máx}\, x_i - min\, x_i$$

Ej: Sea X, las indemnizaciones recibidas por cuatro trabajadores de dos empresas A y B

A	100	120	350	370
B	225	230	240	245

Re (A) = 370 – 100= 270

Re (B) = 245 – 225= 20 ---→ Distribución menos dispersa

- Otros recorridos:

• intervalo intercuartílico $I = Q_3 - Q_1$

• intervalo interdecílico I= $(D_9 - D_1)$

• intervalo intercentílico I= $(P_{99} - P_1)$

b) <u>Desviación absoluta media con respecto a la media (d_e)</u>

Nos indica las desviaciones con respecto a la media con respecto a la media aritmética en valor absoluto.

$$d_e = \frac{\sum_{i=1}^{r} \left| x_i - \bar{x} \right| n_i}{N}$$

c) <u>Varianza</u>

La varianza mide la mayor o menor dispersión de los valores de la variable respecto a la media aritmética. Cuanto mayor sea la varianza mayor dispersión existirá y por tanto menor representatividad tendrá la media aritmética.

La varianza se expresa en las mismas unidades que la variable analizada, pero elevadas al cuadrado.

$$S^2 = \frac{\sum_{i=1}^{r} (x_i - \bar{x})^2 n_i}{N} \qquad S^2 = \frac{\sum_{i=1}^{r} x_i^{\,2} n_i}{N} - \overline{x^2}$$

Propiedades:

1ª) La varianza siempre es mayor o igual que cero y menor que infinito $\left(S^2{}_x \geq 0\right)$

2ª) Si a una variable X la sometemos a un cambio de origen " a " y un cambio de escala " b ", la varianza de la nueva variable Y= a + bX, será:

$$\boxed{\left(S_y{}^2 = b^2 S_x{}^2\right)}$$

d) Desviación típica o estándar

Se define como la raíz cuadrada con signo positivo de la varianza.

$$\boxed{S_x = +\sqrt{S_x^2}}$$

2.4.2. **Medidas de dispersión relativa**

Nos permiten comparar la dispersión de distintas distribuciones.

a) Coeficiente de variación de Pearson (CVx)

Indica la relación existente entre la desviación típica de una muestra y su media.

$$\boxed{CV = \frac{S}{\bar{x}}}$$

Al dividir la desviación típica por la media se convierte en un valor excento de unidad de medida. Si comparamos la dispersión en varios conjuntos de observaciones tendrá menor dispersión aquella que tenga menor coeficiente de variación.

El principal inconveniente, es que al ser un coeficiente inversamente proporcional a la media aritmética, cuando está tome valores cercanos a cero, el coeficiente tenderá a infinito.

Ejemplo: Calcula la varianza, desviación típica y la dispersión relativa de esta distribución.

Sea x el número de habitaciones que tienen los 8 pisos que forman un bloque de vecinos

X	ni
2	2
3	2
5	1
6	3

	N= 8

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i n_i}{N} = \frac{2*2+3*2+5*1+6*3}{8} = 4.125$$ habitaciones

$$S^2 = \frac{\sum_{i=1}^{r} x_i^{\ 2} n_i}{N} - \overline{x^2} = \frac{2^2*2+3^2*2+5^2*1+6^2*3}{8} - (4.125)^2 = 2.86$$ (habitaciones)2

$$S_x = +\sqrt{S_x^2} = +\sqrt{2.86} = 1.69$$ habitaciones

$$CV = \frac{S}{\bar{x}} = \frac{1.69}{4.125} = 0.41$$

2.5. Medidas de forma

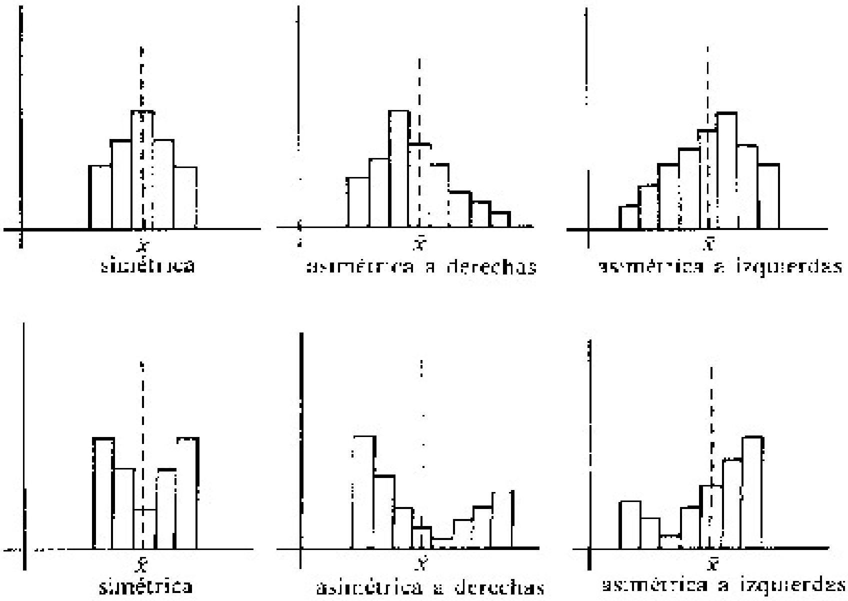
- Asimetría
- Curtosis o apuntamiento.

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.

2.5.1. **Medidas de asimetría o sesgo : Coeficiente de asimetría de Fisher.**

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. En caso contrario, dicha distribución será asimétrica o diremos que presenta asimetría.

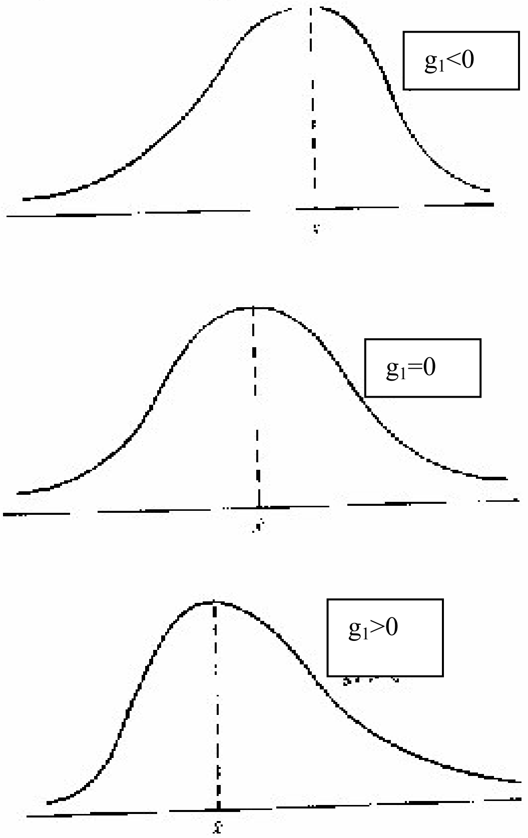


El coeficiente de asimetría más preciso es el de Fisher, que se define por:

$$\boxed{g_1 = \frac{\dfrac{\sum_{i=1}^{r} \left(x_i - \overline{x}\right)^3 n_i}{N}}{S^3}}$$

Según sea el valor de g1, diremos que la distribución es asimétrica a derechas o positiva, a izquierdas o negativa, o simétrica, o sea:

Si g1 > 0 ➔ la distribución será asimétrica positiva o a derechas (desplazada hacia la derecha).
Si g1 < 0 ➔ la distribución será asimétrica negativa o a izquierdas (desplazada hacia la izquierda).
Si g1 = 0 ➔ la distribución puede ser simétrica; si la distribución es simétrica, entonces si podremos afirmar que g1 = 0.



- Si existe simetría, entonces $g_1 = 0$, y $\overline{X} = Me$; si además la distribución es unimodal, también podemos afirmar que: $\overline{X} = Me = Mo$
- Si $g_1 > 0$, entonces : $\overline{X} > Me > Mo$
- Si $g_1 < 0$, entonces : $\overline{X} < Me < Mo$

2.5.2. **Medidas de apuntamiento o curtosis: coeficiente de curtosis de Fisher**

Con estas medidas nos estamos refiriendo al grado de apuntamiento que tiene una distribución; para determinarlo, emplearemos el coeficiente de curtosis de Fisher. (g2)

$$g2 = \frac{\dfrac{\sum_{i=1}^{r}\left(x_i - \overline{x}\right)^4 n_i}{N}}{S^4}$$

Si g2 > 3 la distribución será leptocúrtica o apuntada
Si g2 = 3 la distribución será mesocúrtica o normal
Si g2 < 3 la distribución será platicúrtica o menos apuntada que lo normal.

2.6. Medidas de concentración

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:
x1 = x2 = x3 = ………… = xn-1 = 0 y xn.
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
x1 = x2 = x3 = ………… = xn-1 = xn

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 (xi, ni) de la que formaremos una tabla con las siguientes columnas:

1.- Los productos xi ni, que nos indicarán la renta total percibida por los ni rentistas de renta individual xi .

2.- Las frecuencias absolutas acumuladas Ni .

3.- Los totales acumulados ui que se calculan de la siguiente forma:

u1= x1 n1

u2 = x1 n1 + x2 n2

u3 = x1 n1 + x2 n2 + x3 n3

u4 = x1 n1 + x2 n2 + x3 n3 + x4 n4

un = x1 n1 + x2 n2 + x3 n3 + x4 n4 + ………... + xn nn

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

4.- La columna total de frecuencias acumuladas relativas, que expresaremos en tanto por ciento y que representaremos como pi 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 un y que dada en tanto por ciento, la cual representaremos como qi 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:

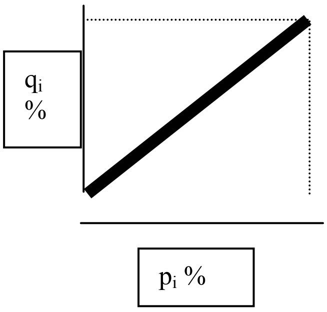
xi	ni	xi ni	Ni	ui	$p_i = \frac{N_i}{n}100$	$q_i = \frac{u_i}{u_n}100$	pi - qi
x1	n1	x1 n1	N1	u1	p1	q1	p1 - q1
x2	n2	x2 n2	N2	u2	p2	q2	p2 - q2
...	...	...	...	...	...	...	...
xn	nn	xn nn	Nn	un	pn	qn	pn - qn

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

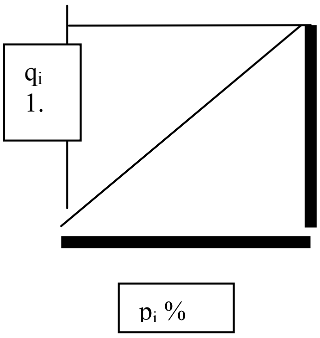
Si esto lo representamos gráficamente obtendremos la curva de concentración o curva de Lorenz .La manera de representarlo será, en el eje de las X, los valores pi en % y en el de las Y los valores de qi en %. Al ser un %, el gráfico siempre será un cuadrado, y la gráfica será una curva que se unirá al 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

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

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

Este índice tomara los valores de I_G = 0 cuando pi = qi concentración mínima y de I_G = 1 cuando qi = 0

Esto lo veremos mejor con un ejemplo :

Li-1 - Li	marca xi	Frecuencia ni	Ni	xini	Σun	qi = (ui/un)* 100	pi=(Ni/n) *100	pi - qi
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

Indice de concentración de GINI

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

, Observamos que hay poca concentración por encontrarse cerca del 0.

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