NōS Oposicións

MATEMÁTICAS

TEMA 60. Parámetros estadísticos. Cálculo, significado y propiedades.

www.nosoposicions.com
www.espazonos.com

ÍNDICE DE CONTIDOS

1. INTRODUCCIÓN

Cuando se hace una estadística de una característica de la población se pretende obtener información cuantitativa de la misma, de modo que podamos aplicar el fenómeno que estamos estudiando. A veces, el número de datos es muy grande, con lo que resulta de difícil interpretación. En estas ocasiones se hace necesario calcular unos indicadores que segmentan (resuman) todos los datos. Así, por ejemplo, la media aritmética se puede usar para resumir (contar) las notas otorgadas por un tribunal a un opositor o se puede usar el rango (diferencia entre el valor máximo y el mínimo) para comparar la "dispersión" de las notas en dos grupos diferentes.

Estos son los dos tipos más comunes de indicadores, los de tendencia central, que pretenden resumir los datos con un valor "central" y los de dispersión que pretenden medir si los datos son muy diferentes entre sí.

Las características que estudiamos en una población pueden clasificarse en dos tipos:

1) Cualitativos: Se describen mediante palabras. Por ejemplo: el color de ojos de una persona.

2) Cuantitativos: Se describen mediante una cantidad. Por ejemplo: el número de hermanos de una persona o su estatura. Estas a su vez se clasifican en dos tipos:

   2.1). Discretas: Sólo pueden tomar una cantidad finita o numerable de valores.

   Por ejemplo: el número de hermanos.

   2.2) Continuas: Pueden tomar todos los valores comprendidos en un intervalo.

   Por ejemplo: la estatura de una persona.

2. DISTRIBUCIONES DE FRECUENCIAS

Una vez obtenida la información (datos) sobre los individuos de nuestra población (por muestreo o por censo) debemos ordenarla y resumirla. Para ello, y especialmente para las características cualitativas o las cuantitativas de tipo discreto, el primer peso suele ser el recuento de las respuestas idénticas.

2.1. Variables discretas

Definición: llamemos "frecuencia absoluta de un dato "al número de veces que se repite dicho dato. Habitualmente a los datos los representariamos por x_1, x_2, x_3,... y a sus frecuencias mediante $f(x_0)$, f_i ó n_i. El número total de datos (individuos) se puede representar mediante n y lógicamente, se verificará

$$n = f_1 + f_2 + \cdots = \sum_i f_i$$

Ejemplo1: Supongamos que en el aula hay 25 alumnos y les preguntamos por el nº de hermanos de cada uno. Supongamos que las respuestas obtenidas se agrupan en la siguiente tabla.

x_0 = "Nº de hermanos"	n_2 = "Frecuencia absoluta del dato X_0"
$x_1 = 0$	$n_1 = 5$
$x_2 = 1$	$n_2 = 10$
$x_3 = 2$	$n_3 = 4$
$x_4 = 3$	$n_4 = 3$
$x_5 = 4$	$n_5 = 3$
	TOTAL n = 25

Definición: Llamamos "frecuencia relativa de un dato" al variante entre su frecuencia absoluta y el número total de individuos. La frecuencia relativa de x_i se representa mediante h_i.

2.2. Variables continuas

En este tipo de variables, los datos son muy numerosos y se repiten poco, por lo que contar las frecuencias de cada dato no produce ninguna ventaja a la hora de estudiar y sacar consecuencias de los mismos. Por ejemplo, si estudiamos el tamaño de los mejillones gallegos y tenemos una muestra de 10 de ellos, podemos obtener:

Ejemplo 2: 2'9 cm, 3'4 cm, 2'92 cm, 4 cm, 3'82 cm,
3'12 cm, 3'7 cm, 4'1 cm 3'82 cm, 3'6 cm

Evidentemente, las frecuencias para la mayoría de los datos son 1, por lo que las tablas de frecuencia aportan para información. Por eso, conviene agrupar los datos en intervalos (clases) y calcular las frecuencias para dichos intervalos. Por ejemplo, los datos anteriores se pueden resumir en la siguiente tabla de frecuencias:

Ejemplos:

Intervalo	n
[2'5, 3)	2
[3, 3'5)	2
[3'5, 4)	4
[4, 4'5)	2
	Total n = 10

Definición: Llamaremos "amplitud del intervalo" o "amplitud de la clase" (li ,(li+1)) a su longitud (medida) $a_i = l_{i+1} - l_i$. La amplitud de los intervalos puede ser distinta, aunque es recomendable que todos tengan la misma. Llamaremos "marca de clase" de un intervalo (li ,(li+1)) al punto medio $\frac{l_i+l_{i+1}}{2}$. Cuando los intervalos sean infinitos, la media de clase se calculará de forma análoga a la de los correspondientes intervalos finitos. Así, en el ejemplo anterior si el último intervalo fuese (4 ,(∞)), la marca de clase seguiría siendo 4'75.

<u>Definición</u>: Llamaremos "recorrido" o rango" de una variable a la diferencia entre el valor máximo y mínimo $R = x_{max} - x_{min}$

En el ejemplo 2: $a_i = 0{,}5, \forall i = 1{,}2{,}3{,}4$

$$x_1 = \frac{2{,}5 + 3}{2} = 2{,}75$$
$$x_1 = \frac{3 + 3{,}5}{2} = 3{,}25$$
$$x_1 = \frac{3{,}5 + 4}{2} = 3{,}75$$
$$x_1 = \frac{4 + 4{,}5}{2} = 4{,}25$$
$$R = x_{max} - x_{min} = 4{,}1 - 2{,}9 = 1{,}2$$

Cuando agregamos los datos en intervalos se nos plantea el problema de cuántos intervalos debemos construir y cómo. Aunque la agrupación se puede hacer como uno quiera, ésta sólo será efectiva si el número de intervalos es el adecuado, ya que tanto un número demasiado grande como uno demasiado pequeño provocarán que las correspondientes tabla de frecuencias sean para informativos. Por ello, hay que llegar a un "compromiso" entre un único intervalo (perdemos toda la información de los datos) y tantos intervalos como datos (el exceso de información no permite sacar consecuencias).

Para determinar el número de intervalos existen numerosas reglas. La más popular es la regla de Sturges (Kelley) que establece que si tenemos n datos el número de intervalos debe ser:

$$k \cong 1 + \log_2 n$$

Como K debe ser entero, debemos redondear el entero más próximo.

Una vez determinado el número de intervalos, si queremos que todos ellos tengan la misma amplitud, ésta se calculará mediante:

$$a(amplitud) = \frac{R(rango)}{k(nº\ de\ intervalos)}$$

Y los intervalos se construyen a partir del valor mínimo (o máximo), sumando (restando) la amplitud:

$(l\,1\,,\,(l\,2\,)) = (\,xmi\,n\,,\,(xmi\,n + a\,))$
$(l\,2\,,(l\,3\,)) = (\,xmi\,n\,,\,(xmi\,n + 2\,a\,))$

Haciendo que el último intervalo (primero) sea cerrado. Si la amplitud se redondea, debe hacerse hacia el mayor valor, o serán necesarios $k+1$ intervalos para cubrir a los datos.

EJEMPLO 2: $\begin{pmatrix} k \simeq 1 + \mathrm{lo}\ \ _{10}\, 2 \\ a = \frac{R}{k} = 0{,}277\ldots \end{pmatrix}$

3. MEDIDAS DE TENDENCIA CENTRAL

Veamos las medidas de tendencia central que se utilizan para resumir el conjunto de los datos con un único valor.

3.1. Media aritmética

El indicador de tendencia central más utilizada es la "media aritmética" (o simplemente media), que representaremos mediante $\overline{x}$ y viene definida por:

$$\overline{x}=\frac{n_1.x_1+n_2.x_2+\cdots+n_k.x_k}{n}=\frac{1}{n}\sum_i n_i.x_i\sum_i h_i.x_i$$

Cuando los datos provengan de una muestra de la población total suele denominarse "media muestral"

Ejemplo 1: La media viene dada por: $\overline{x}=\frac{0\cdot5+1\cdot10+2\cdot4+3\cdot3+43}{25}=1.,16$

Si la variable en estudio es continua y los datos están agrupados en intervalos, para calcular la media usaremos las marcas de clase x_i y las frecuencias de cada clase n_i. Por ejemplo:

Ejemplo 2: La media viene dada por: $\overline{x}=\frac{2{,}75\cdot2+3{,}25\cdot2+3{,}75\cdot4+4{,}25\cdot2}{25}=3{,}45$

Para el cálculo de la media puede utilizarse la siguiente tabla:

Intervalo	x_i	n_i	$x_i\cdot n_i$
[2'5, 3)	2'75	2	5'5
[3, 3'5]	3'25	2	6'5
[3'5, 4)	3'75	4	15
[4, 4'5)	4'25	2	8'5
TOTAL		10	34'5

$$\overline{x}=\frac{34'5}{10}=3'45$$

En ocasiones es conveniente que algunos datos tengan más importancia que otros, en ese uso podemos usar la "media aritmética ponderada "definida por:

$$m_{\omega(X)}=\frac{\sum_i \omega_i}{\sum_i \omega_i\cdot x_i}$$

Donde los ω_i: son números positivos. Por ejemplo, las calificaciones de tres ejercicios de una oposición se puedan ponderar con 1,15 y 2.

Propiedades de la media aritmética:

Denotamos por m (x) la media de una variable estadística x, se verifican las siguientes propiedades.

1) $m(ax+b) = am(x) + b, \forall a, b \in R$

Demostración: $m(ax+b) = \frac{1}{n}\sum_i (ax_i + b)n_i = \frac{1}{n}\sum_i (ax_i n_i) + \frac{1}{n}\sum_i (bn_i) = a\frac{1}{n}\sum_i (x_i n_i) + \frac{1}{n}bn = am(x) + b$

2) La media es la única solución de la ecuación (en z):

$$\sum_i (x_i - z)n_i = 0$$

Es decir, el número que, tomando como origen de coordenadas, hace que la suma de las distancias (con signo) al cero de todos los datos sea cero (centro de gravedad), es la media aritmética. La demostración es inmediata, ya que

$$\sum_i (x_i - z)n_i = 0 \Rightarrow \sum_i x_i n_i - z\sum_i n_i = 0 \Rightarrow \sum_i x_i n_i - zn = 0 \Rightarrow z = \frac{\sum_i x_i n_i}{n} = m(x)$$

3) La suma de las diferencias al cuadrado de todos los datos con $m(x) = \bar{x}$ es mínima, es decir: la única solución de $\min q(z) = min(\sum_i (x_i - z)^2 n_i)$ es $z = m(x) = \bar{x}$.

Demostración: $q(z) = \sum_i (x_i - z)^2 n_i$ es derivable respecto a z, por ser una función polinómica. Además, $q'(z) = -2\sum_i (x_i - z)\, n_i$. en consecuencia:

$$q'(z) = 0 \Rightarrow \sum_i (x_i - z)\, n_i = 0 \Rightarrow z = \bar{x}$$

Y como $q''(z) = 2\sum_i n_i = 2n > 0$ se tiene que $\bar{x}$ es un mínimo.

4) Si unimos varios conjuntos de datos, la media conjunta será igual a las correspondientes medias aritméticas de cada conjunto, con pesos iguales a la cantidad de datos de cada conjunto. Es decir, si disponemos de los datos $\{x_{11}, x_{12}, \dots, x_{1n(1)}\}, \{x_{21}, x_{22}, \dots, x_{2n(2)}\}, \dots, \{x_{k1}, x_{k2}, \dots, x_{kn(k)}\}$

con sus correspondientes frecuencias n_{ij} entonces:

$\bar{x} = \frac{\sum_i n(i)\bar{x}_i}{\sum_i n(i)}$ donde $\bar{x}_i$ es la media de los datos del grupo i-'esimo, es decir: $\bar{x}_i = \frac{\sum_j x_{ij} n_{ij}}{n(i)}$ con $x_{ij} = x_{in(j)}$ y n_{ij} es la frecuencia absoluta de x_{ij}.

Demostración:

$$\frac{\sum_i n(i)\bar{x}_i}{\sum_i n(i)} = \frac{\sum_i n(i)\bar{x}_i}{n} = \frac{\sum_i n(i)\left(\frac{\sum_j x_{ij} n_{ij}}{n(i)}\right)}{n} = \frac{\sum_i \sum_j x_{ij} n_{ij}}{n} = \frac{\sum_k x_k n_k}{n} = \bar{x}$$

5) La media es "recursiva", es decir si añadimos un nuevo dato, la media x puede recalcular a partir de la media anterior, el número de datos y el nuevo dato. Además, la fórmula de recursión es :

$$m(x_1, x_2, \dots, x_{n+1})=\frac{n}{n+1}\, m(x_1, x_2, \dots, x_n) + \frac{1}{n+1} x_{n+1}$$

La demostración es consecuencia de 4 haciendo dos grupos, uno con la n primeros datos y otro con "el nuevo":

Ventaja de la media aritmética:

- Es servible a la variación en cualquier dato. (Basta con que centro un dato para que cambie su media).
- Si cambiamos las unidades es que se mide la variable, la media cambia de igual forma.
- Tiene una fácil integración (centro de gravedad, mínima distancias al cuadrado, etc) y un cálculo sencillo.
- Es única y siempre existe.
- Es la medida de centralización más utilizada, por lo que aparece en calculadoras y programas del ordenador.

Inconvenientes:

- Los valores extremos influyen mucho en el valor de la media por lo que los datos erróneos o muy dispares pueden hacer que no se representativa del conjunto de datos.
- Al ser sensible: al cambio de unidades, no se puede utilizar para comparar variables que se midan es unidades diferentes.
- Depende de los intervalos (clases) y de las marcas de clase, por lo que se puede prestar a manipulaciones.

3.2. Media geométrica

Se define mediante $m_g = \sqrt[n]{{x_1}^{n_1} \cdot {x_2}^{n_2} \cdot \cdots \cdot {x_k}^{n_k}} = e^{\frac{1}{n}\sum_i n_i \log x_i}$ Por los tanto, será la media aritmética de los datos$z_i = \log x_i$. De nuevo, si los datos están agrupados en intervalos, para calcular la media utilizaremos las marcas de clase x_i.

Propiedades de la media geométrica:

1) Si combinamos la escala (multiplicamos por una constante) de los datos, la media cambia en la misma medida, es decir:

$$m_g(ax) = a m_g(x)$$

2) Si elevamos los datos a cualquier potencia (constante), la media geométrica cambia en la misma medida.

$$m_g(x^b) = m_g(x)^b$$

3) La media geométrica es la única solución de la ecuación (en z):

$$\sum_i (\log x_i - \log z) n_i = 0$$

4) La media geométrica minimiza la expresión $mín\, q(z) = mín \sum_i (\log x_i - \log z)^2\, n_i$

5) Si unimos varios conjuntos de datos, la media geométrica conjunta será igual ….. La media geométrica ponderada de las correspondientes medias geométricas de cada conjunto, con pesos iguales a la cantidad de datos de cada conjunto.

6) La media geométrica es "recursiva". La fórmula de recursión es:

$$m_g(x_1, x_2, \ldots, x_{n+1}) = \sqrt[n+1]{x_{n+1} m_g(x_1, x_2, \ldots, x_{n+1})^n}$$

Ventajas de la media geométrica:

- Es sensible a la variación en cualquier dato.
- Si cambiamos la escala o elevamos a una potencia (cambia de base) las unidades es que se mida la variable, la media cambia de igual forma.
- Es única.
- Los valores extremos influyen poco.
- Es sensible de calcular.
- Suele utilizarse para resumir razones (cocientes), porcentaje, etc.

Inconveniente:

- No siempre existe. De hecho, puede no ser posible su cálculo si algún x_i es negativo o no tiene mucho sentido si algún x_i es 0.
- Los valores cercanos a 0 influyen mucho en su valor.
- No se puede utilizar para comparar variables que se pidan en unidades diferentes.
- Depende de los intervalos y de las marcas de clase, por lo que se puede prestar a manipulaciones.
- Su significado es menos intuitivo.
- No suele incluirse en calculadora y programas de ordenador.

3.3. Mediana

Tras la media aritmética, el indicador de tendencia central más utilizado es la media que se define como el valor que deje un 50% de datos a su izquierda y un 50% a su derecha.

$$\frac{1}{n}\sum_i 1(x_i \le me) \ge 0'5$$
$$\frac{1}{n}\sum_i 1(x_i \ge me) \ge 0'5$$

Nótese que, si calculamos las frecuencias relativas acumuladas Nº a la izquierda y a la derecha, la mediana es el valor que verifica simultáneamente que ambas son mayores que 0'5. Si ordenamos los datos, la mediana sería el valor que origine el lugar central (si el número de datos es impar) o cualquier valor situado entre los datos centrales (si el número de datos es par). Para que la mediana tenga un valor único, otros autores prefieran definirla para un número par de datos, como la media aritmética de los dos valores centrales. Una forma de calcularla es ordenar los datos. Por ejemplo:

0,0,0,0,,1,1,1,1,1,1,1,1,1,1,2,2,2,2,3,3,3,4,4,4

Entonces me = 1

También se puede calcular mediante la tabla de frecuencias acumuladas. Para ello, buscaremos el valor más pequeño de los x_1 que tiene frecuencia acumulada N_i mayor o igual que n/2. : Si $N_i > \frac{n}{2}$, entonces $me = x_i$ .Si $N_i = \frac{n}{2}$ entonces $me = \frac{x_i + x_{i+1}}{2}$

Si los datos están agrupados en intervalos se procede de forma similar. En primer lugar, obtenemos la primera clase (li ,(li+1)) para la que se verifica $N_i \geq \frac{n}{2}$ y entonces la mediana se define como: $me = l_i + \frac{\frac{n}{2} - N_{i-1}}{n_i}(l_{i+1} - l_i)$.

Propiedades de la mediana:

1) $me(ax + b) = a \cdot me(x) + b$
2) La mediana minimiza la expresión: $\sum_i (x_i - z)$
3) La mediana es el puesto que divide al histograma en dos partes iguales.
4) Si unimos varios conjuntos de datos, la mediana no es, en general, la media ponderada del conjunto de las medianas.
5) La mediana no es necesaria.

Ventajas de la mediana:

- Para su cálculo se usan todos los datos, pero es menos sensible que la media a la variación de cualquier dato.
- Si cambiamos la escala de las unidades en que se mide la variable, la media cambia de igual forma.
- Es única y siempre existe.
- Es más recomendable que la media cuando la distribución es asimétrica.
- Casi nunca se ve afectada por la forma de definir los intervalos extremos si estos son indefinidos.
- Se suele incluir en casi todos los programas de estadística para ordenar.

Inconvenientes:

- No es tan sencilla de calcular como la media.
- Al ser sensible al cambio de escala en las unidades, no se puede utilizar para comparar variables que se midan en unidades diferentes.

3.4. Moda

Tras la media y la mediana es el indicado de tendencia central más utilizado. Se define como el valor con mayor frecuencia absoluta.

$$mo = x_j \,;\; n_j \geq n_i \,, \forall i$$

Es el dato que más se repite. Puede que existan varios puntos que tengan frecuencia máxima, por lo que la moda no es única. Su cálculo a partir de la tabla de frecuencias absolutas es muy sencillo. Los datos están agrupados en intervalos, la moda se define como la media de clase de clase del intervalo con mayor frecuencia absoluta.

Propiedad de la moda:

1) $mo(ax+b) = a \cdot mo(x) + b$
2) Si unimos varios conjuntos de datos, la moda no es, en general, la moda ponderada del conjunto de las modas.
3) La moda no es recursiva.

Ventajas de la moda

- Es sencilla de calcular.
- Si cambiamos la escala es que se midan las unidades, la moda cambia de igual forma.
- Tiene una interpretación sencilla y siempre existe.
- Los valores extremos prácticamente no influyen en ella.
- Se suele incluir en casi todos los programas de estadística para ordenador.

Inconvenientes de la moda.

- Para su cálculo no se usan todos los datos
- En general, no es única.
- Al ser sensible al cambio de escala en las unidades, no se puede utilizar para comparar variables medianas en unidades diferentes.
- Depende de los intervalos (clases) y puede estar sujeta a manipulaciones.

3.5. Otras medidas de posición

- Media armónica: $m_a = \frac{n}{\sum_i n_i {x_i}^{-1}}$
- Media cuadrática: $m_c = \sqrt{\frac{\sum_i n_i {x_i}^2}{n}}$
- Aunque no son medidas de centralización, sí son medidas de posición los "cuartiles" y "percentiles".

4. MEDIDAS DE DISPERSIÓN

Tras obtener una medida de centralización podemos estudiar si ésta es representativa del conjunto de datos midiendo la dispersión de los datos. Veamos las medidas de dispersión más utilizadas:

4.1. Varianza y cuasivarianza

Los indicadores de dispersión más utilizados son la "varianza" (s^2) y la "cuasivarianza" ($S^{"}$) que se definen mediante:

$$s^2 = Var(X) = \frac{1}{n}\sum_i (x_i - \bar{x})^2\, n_i$$

$$S^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2\, n_i$$

Evidentemente $S^2 = \frac{ns^2}{n-1}$.Para calcular la varianza, se usa la relación:

$$s^2 = \frac{1}{n}\sum_i {x_i}^2 n_i - \bar{x}^2$$

Ejemplo 2:

Intervalo	x_i	n_i	$x_i n_i$	$x_i {n_i}^2$
[2'5, 3)	2'75	2	5'5	11
[3, 3'5)	3'25	2	6'5	13
[3'5, 4)	3'75	4	15	60
[4, 4'5)	4'25	2	8'5	17
TOTAL		10	35'5	101

$$s^2 = \frac{101}{10} - 3'5^2 = 8'126$$

Propiedades:

1) $Var(ax+b) = a^2 Var(X)$
2) La varianza es el mínimo de las medias de las desviaciones cuadráticas respecto a cualquier punto.

$$s^2 = \min_z \frac{1}{n}\sum_i (x_i - z)^2$$

3) La varianza es necesaria.

4) Si (x_i, n_i) es una distribución de frecuencias con media $\bar{x}$ y varianza s^2y t > 0, entonces el intervalo $(\bar{x} - ts, \bar{x} + ts)$ contiene a más del $100 \cdot \left(1 - \frac{1}{t^2}\right)$ % de los datos.

Ventajas:

- Es por cambios de localización.
- Siempre existe y tiene interpretación sencilla.
- Para su cálculo se usan todos los datos.
- Junto con la media, nos da una idea de dónde están la mayoría de los datos.
- Se suele incluir en calculadoras científicas y en programas de estadística.

Inconvenientes:

- No es invariante por cambio de escala y no se expresa en las mismas unidades que los datos.
- Los valores extremos influyen en ella.
- Su cálculo es complicado.
- No se puede usar para comparar variables medidas en unidades diferentes.
- Depende de los intervalos, por lo que se presta a manipulaciones.

4.2. Desviación estándar y cuasidesviación estándar

El principal problema de la varianza y de la cuasivarianza es que no se expresa en las mismas unidades que los datos. Por ello, a veces es preferible utilizar la ”desviación estándar (s)” o la ”cuasidesviación estándar” (S), que se definen mediante:

$$s = \sqrt{\frac{1}{n}\sum_i (x_i - \bar{x})^2\, n_i}$$

$$S = \sqrt{\frac{1}{n-1}\sum_i (x_i - \bar{x})^2\, n_i}$$

Evidentemente $S = \sqrt{\frac{n}{n-1}}\, s$ y su cálculo es similar al de la varianza.

Propiedades.

1) $DS(ax + b) = aDS(x)$

La desviación estándar es inconveniente por centro de localización y cambia de igual forma al hacer cambios de escala.

2) La desviación estándar es el mínimo de la raíz cuadrada de la media de las desviaciones cuadráticas de cualquier puesto.

$$s = \min_z \sqrt{\frac{1}{n}\sum_i (x_i - z)^2}$$

3) La desviación estándar es recursiva.

Ventajas:

- Se miden en la misma escala que los datos.
- Es invariante por cambios de localización.
- Tiene interpretación sencilla y siempre existe.
- Para su cálculo se usan todos los datos.
- Junto con la media nos da una idea de dónde se encuentran la mayoría de los datos.

Inconvenientes:

- Los valores extremos influyen mucho en ella.
- Su cálculo es algo complicado.
- Al ser sensible al cambio de escala en unidades, no se puede usar para comparar variables que se midan en unidades diferentes.
- Depende de los intervalos (clases), por lo que se puede prestar a manipularse.

4.3. Otras medidas de dispersión

1) Rango o recorrido: Es la diferencia entre el máximo y mínimo de los datos.

$$R = \max_i x_i - \min_i x_i$$

2) Desviación media: $DM = \frac{1}{n}\sum_i |x_i - \bar{x}|$

3) Desviación mediana: $DMe = \frac{1}{n}\sum_i |x_i - me|$

4) Coeficiente de variación de

   Pearson: $CV = \frac{s}{\bar{x}}$, si $\bar{x} \neq 0$.

5. MEDIDAS DE FORMA

En otras ocasiones podemos estar interesados en la forma, simetría o "aplastamiento" de los datos [curtosis).

1) Momentos: Son medidas de forma.

$$m_k = \frac{1}{n}\sum_i (x_i - \bar{x})^k$$

2) Coeficiente de asimetría de Fisher:

$$g_1 = \frac{m_3}{s^3}$$

Si $g_1 > 0$ la representación de los datos presenta asimetría a la derecha y en cambio si $g_1 < 0$ la asimetría se presenta a la izquierda.

3) Coeficiente de aplastamiento: $g_2 = \frac{m_4}{s^4} - 3$

6. BIBLIOGRAFÍA

- Engel, A.: ”Probabilidad y Estadística”.
- Gmurman, V. E.: ”Teoría de las probabilidades y estadística matemática”.
- Mood, A. Graybill, F. Boes, D.: ”Introduction to the theory of statistics”.RUBIO ”.
- Renyi, A: ”Calcul des probabilities”.
- Ríos S. : ”Análisis estadístico aplicado”.