MEDIA, DESVIACIÓN TÍPICA Y VARIANZA

Cuando disponemos de una gran masa de datos y queremos hacernos una idea de ellos en su conjunto, deberemos calcular tres parámetros:

1) $n \equiv$ número de datos.
2) Un valor que esté en el "centro" de los datos.
3) Un valor que nos diga cómo están los datos distribuidos respecto del "centro", o sea, si están muy cerca del mismo, o muy dispersos.

Supongamos que tenemos n datos: $x_1, x_2, \ldots, x_n$. Veamos cómo obtener dichos parámetros.

Media

El valor más útil como "centro" de los datos, desde el punto de vista de propiedades matemáticas, es el conocido como *media aritmética*. Cuando el número de datos n es finito, se obtiene sumándolos y dividiendo el resultado de dicha suma entre n:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

donde $\sum_{i=1}^{n} x_i$ es una manera corta de escribir una fórmula compuesta de n sumandos que tienen todos ellos la misma estructura. Se denomina *sumatorio*, x_i es la estructura de cada sumando (en este caso), $i = 1$ nos proporciona el primer sumando, $i = n$, el último, aumentando i de uno en uno.

<u>Ejemplo</u>: Si disponemos de los siguientes datos:

2 3 5 5 7 9 9 9 10 12

La *media aritmética* de los mismos valdrá:

$$\overline{x} = \frac{2+3+5+5+7+9+9+9+10+12}{10} = \frac{2+3+5\cdot 2+7+9\cdot 3+10+12}{10} = 7.1$$

Observar que cuando hay datos repetidos, en lugar de repetir un mismo sumando varias veces, es más cómodo usar un único sumando con el dato multiplicado por el número de veces que se repite (a lo que se llama *frecuencia* del dato).

Desviación típica y Varianza

El valor más útil, desde el punto de vista de sus propiedades matemáticas, como medida de la desviación de los datos respecto del valor central de los mismos, se llama *desviación típica*. Cuando tenemos n datos, hay dos formas de calcularlo (la segunda es más fácil):

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2}$$

que, sin *sumatorios* es:

$$s = \sqrt{\frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \ldots + (x_n - \overline{x})^2}{n}} = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n} - \overline{x}^2}$$

A menudo, se usa también el cuadrado de la desviación típica, lo que se conoce como *varianza*. Esta última puede ser más cómoda, en ocasiones, porque no hay que calcular la raíz cuadrada final; pero la *desviación típica* se mide en las mismas unidades que los datos, mientras que la *varianza* se mide en dichos datos *al cuadrado*. O sea, si los datos son, por ejemplo, *pesos en kg*, la *media* y la *desviación típica* están en *kg*, pero la *varianza*, en *kg*^{2}, por lo que la información que nos da, a primera vista, no es intuitiva.

Ejemplo: Para los mismos datos del ejemplo anterior, la *varianza* valdrá:

$$s^2 = \frac{2^2+3^2+5^2+5^2+7^2+9^2+9^2+9^2+10^2+12^2}{10} - 7.1^2 =$$

$$= \frac{2^2+3^2+5^2\cdot 2+7^2+9^2\cdot 3+10^2+12^2}{10} - 7.1^2 = 9.49$$

Por tanto, la *desviación típica* es: s = 3.0806, que es la raíz cuadrada de la *varianza*.

Las fórmulas de *media* y *desviación típica* con las frecuencias

Como hemos visto en el ejemplo cuando alguno de los datos está repetido es más fácil ponerlo una vez y multiplicarlo por el número de veces que se repite, que escribir tantos sumandos como repeticiones. Al número de veces que un dato aparece repetido se le llama *frecuencia*. Y, tal como se ha hecho en los ejemplos, la *media* (que, escrita así, se llama, a veces, *media ponderada*) y la *desviación típica* con *frecuencias* son:

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i f_i}{n} \qquad s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 f_i}{n} - \overline{x}^2}$$

Uso de la calculadora

Cuando resolvemos un problema, debemos siempre indicar de dónde obtenemos los datos. Por tanto, hay que escribir las fórmulas anteriores. Pero el cálculo de sus resultados es más cómodo obtenerlo con la calculadora en *modo estadística*. No nos es posible indicar cómo hacerlo para todos los modelos, porque, por desgracia, los fabricantes lo cambian de un modelo a otro. Por ello, aconsejamos consultar los manuales. A modo de ejemplo, comentamos el de algunos modelos tipo Casio.

Lo primero, es colocar la calculadora en *modo estadística de 1 variable*. Suele hacerse con algo parecido a *Mode* (puede que haya que pulsarlo más de una vez), y elegir *SD* ó *1-Var*. Es fundamental recordar volver al modo normal una vez terminado, lo que se hace de la misma forma, eligiendo *COMP*, porque de lo contrario hay opciones usuales de la calculadora que no estarán disponibles. Si la calculadora muestra una tabla de una columna para que introduzcamos los datos, entrar en *Setup* o *Config*, y elegir en *SD*, *1-Var* o *Stat* la opción *Frequency ON*, para que aparezca una segunda columna con las frecuencias.

El paso siguiente es introducir los datos. Si es en modo tabla, como hemos descrito antes, no hay más que colocarse en la posición adecuada, escribir el número y pulsar la tecla de introducción (normalmente, la tecla =), moviéndonos con la tecla redonda central con flechas para introducir más datos. Si nuestra calculadora no funciona en modo tabla, escribimos el *dato*, punto y coma (*Shift ,*) seguido de la *frecuencia* y pulsamos *M+*.

Por último, las teclas *S-SUM* y *S-VAR* o similares, o bien las opciones *Data* o *Stat* nos proporcionan los resultados (*n*, *media* y *desviación típica*). Hay que señalar que las calculadoras muestran dos opciones para la *desviación típica*, señaladas con *s* ó σ con algún subíndice. La primera opción mostrada es la *desviación típica* y la segunda, la *cuasidesviación típica*, cuya fórmula es idéntica a la de la *desviación típica* pero con denominador $n - 1$ en lugar de n.

Muestra y Población

Al estudiar la *Población* al completo, puede que ésta sea infinita. Para la *población*, la *media* y la *desviación típica* serían teóricas, y su cálculo es una generalización de estas fórmulas con instrumentos matemáticos más poderosos. En el caso de la *población*, la *media* se llama también *esperanza matemática* o *valor esperado*, y se designa por μ, y la *desviación típica*, por σ.