CienciaOnline
$\int$ Matemáticas $\int$

Lorenzo Hernández

3 de diciembre de 2013

La media no está en el medio. $\int$

Se suele decir que la estadística es la manera más fácil de engañar a alguien (que no sabe estadística, añado yo). Uno de los conceptos más sencillos de entender y que todo el mundo ha estudiado es la media aritmética. Por sencillo que resulte, y aunque todo el mundo la haya estudiado en matemáticas, es sorprendente la cantidad de titulares de prensa y ruedas de prensa de políticos que usan este concepto para engañar y ocultar información. Parece que, aunque la hayamos estudiado, quizá la mala memoria, junto al propio término (media) que nos puede despistar, hayamos olvidado (sobre todo si eres político) una idea sencilla:

"La media no está en el medio."

Es decir, en un grupo de números, no tiene que haber la misma cantidad de números a un lado que a otro de la media. Una de las limitaciones de la media aritmética es que se trata de una medida muy sensible a los valores extremos; valores muy grandes tienden a aumentarla mientras que valores muy pequeños tienden a reducirla, lo que implica que puede dejar de ser representativa de la población.

Hay que saber varias cosas de la media:

- Que nos ocultan todo el batiburrillo del mundo real.
- Que pueden ser una excepción y no la norma. Pueden parecer lo corriente, cuando quizá se definan por ser extraordinarias, hallándose lejos de la mitad.

Resulta curioso que en el lenguaje diario, “medio” es una palabra que significa bajo o pequeño (del montón) pero cuando lo aplicamos a ingresos, sin embargo, la media es relativamente alta.

Si hablamos de números quizás no nos importe demasiado, pero si hablamos del sueldo medio en España quizá le prestaremos más atención. Esto se puede ilustrar con un ejemplo muy sencillo:

Imaginemos que un profesor dice que los resultados de su grupo de estudiantes han sido buenos, ya que la media de la nota académica es de un 5,2 (aceptarme esta media como buena). Pero la realidad es que, si suponemos que tiene 10 estudiantes, dos de ellos han obtenido un 10 y los otros ocho han obtenido un 4, es decir, sólo dos han superado la asignatura. “La media no está en el medio”, ya que hay dos a un lado y ocho al otro lado.

Lo mismo ocurre con lo beneficios fiscales, las subidas de impuestos, la bajada de salarios, los precios medios de la vivienda, la esperanza de vida y un largo etcétera. Las medias lo mezclan todo. De ahí que sean útiles, pero debemos ir con pies de plomo cuando nos ofrecen una media como resumen. En el sueldo medio se puede meter desde el barrendero hasta el magnate de fondos de inversión que colecciona yates.

Si lo aplicamos al típico titular:

"El salario medio en España es de 22790 euros"

Podemos deducir rápidamente que pueden haber pocas personas que cobren mucho más de 22790 y muchas más que cobren menos, de tal modo que los pocos ricos suben la media de manera que da la sensación de que los sueldos en España son aceptables.

Como también sabréis, la estadística tiene otras herramientas para destripar los datos. Por ejemplo, en un artículo del idealista.com, tras el titular "El salario medio en España es de 22790 euros", continua dando datos que nos reflejan la realidad de la situación. La moda, el salario más frecuente es de 16490 euros y la mediana, otro tipo de media que divide en dos la población, es de 19017 euros, es decir, la mitad de la población cobra más de este nivel y la otra mitad cobra más. Esto significa que hay pocos trabajadores con salarios muy altos pero que influyen notablemente en el salario medio.

A veces hay que usar la media porque, a pesar de todos sus riesgos, necesitamos una cifra que hable en nombre del grupo. Pero que sea útil depende sobre todo de que sepamos de antemano qué grupo es el que nos interesa. Si contemplamos las medias sin saber de dónde las hemos abstraído, nos llevará a confusión.

Lo que suelen ocultar los políticos es cómo han obtenido esas medias y qué datos han usado para calcular la media. En estos días de noticias económicas diarias no está de más recordar estas nociones de estadística básica para estar alerta cuando escuchamos la palabra "media".

Un récord armónico. $\int$

Cuando se supera un récord debemos de tener en cuenta si ha sido independiente o dependiente de otros récords anteriores. Por ejemplo, en las olimpiadas los récords que se baten son dependientes de los anteriores. Cada nueva marca es el resultado de un esfuerzo competitivo que no es independiente de todos los intentos anteriores de realizar la misma proeza. Los saltadores de pértiga aprenden nuevas técnicas, se mejoran los trajes de los nadadores, etc.

Existen, sin embargo, otra clase de récords que surgen en secuencias de acontecimientos que, se da por hecho, son independientes entre sí. Un buen ejemplo es el récord de temperatura en la Tierra.

¿Podremos predecir el siguiente récord? ¿Nos pueden decir las matemáticas algo sobre si hay o no calentamiento global o, al menos, que la subida de temperatura no es aleatoria sino que forma parte de una tendencia sistemática no aleatoria? Esto es lo que se pregunta John D. Barrow en uno de los capítulos de su libro «El salto del tigre. Las matemáticas de la vida cotidiana». Barrow argumenta lo siguiente.

Supongamos que queremos medir la temperatura una serie de años y contar los récords que se producen. El primer año en que empezamos a medir la temperatura contará como el primer récord. En el segundo año, si la temperatura es independiente de la del año anterior, la probabilidad de que supere el registro del primer año es de $1/2$. Por tanto, la cantidad de años récord que esperamos en los dos primeros años es $1+1/2$. En el tercer año hay apenas dos formas en que las seis posibles clasificaciones (esto es, una posibilidad de uno entre tres) de las temperaturas de los años 1, 2 y 3 pueden producir un récord en este último año. Por tanto, el número de años récord después de tres años de registros es $1 + 1/2 + 1/3$. Si seguimos adelante se hallará que después de n años independientes de recopilación de datos, la cantidad esperada de años récord es la suma de una serie de n fracciones:

$$1+ \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \ldots \frac{1}{n}$$

Ésta es la famosa serie "armónica". Si llamamos H(n) a la suma total de n términos, tenemos que H(1) = 1; H(2) = 1,5; H(3) = 1,833; H(4) = 2,083; y así sucesivamente. Los más interesante acerca de la suma de esta serie es que crece de forma muy lenta a medida que el número de términos aumenta, de manera que H (256) = 6,12, pero H (1000) = 7,49 y H (1000000) = 14, 39.

De hecho, cuando n se hace muy grade H(n) aumenta sólo tan rápido como el logaritmo de n y se aproxima muchísimo a 0,58 + ln(n).

¿Nos dice esto algo sobre los récords de temperatura y una posible subida de las temperaturas? Nos indica que los récords, cuando los acontecimientos se producen de forma aleatoria, son muy raros.

Supongamos que hubiéramos podido medir la temperatura en España desde 1748 hasta el 2004, un período de 256 años. Entonces predecimos que deberíamos de hallar sólo H (256) = 6,12, o aproximadamente seis récord de máximas (o mínimas) temperaturas y tendríamos que esperar más de un millar de años para tener posibilidades favorables de hallar siquiera ocho récord.

Si los nuevos récord se convierten en bastantes más comunes de lo que la serie armónica predice, significaría que los sucesos climáticos anuales ya no son acontecimientos independientes, sino que están empezando a formar parte de una tendencia sistemática no aleatoria.