

¡Muera la media, viva la mediana!

Todo el mundo conoce el chascarrillo que suele hacerse para burlarse de la validez de las estadísticas: “La estadística es esa ciencia que cuando yo he comido dos pollos y tú no has comido ninguno, dice que hemos comido uno cada uno”. Como broma, es ingeniosa. Como crítica, es endeble por varias razones. La primera, claro, es que la estadística no dice tal cosa. La media no es ni pretende ser una descripción de lo que sucede a cada uno de los sujetos estudiados, que lógicamente

tendrán valores distintos. Ni siquiera aspira a dar un dato que sea, de manera aproximada, similar al de la mayoría de los casos de la población. Nada hay en su fórmula de cálculo (la suma de todos los valores de todos los casos, dividida entre el número de casos) que garantice que el resultado sea un valor común, reconocible por la mayoría de los miembros del grupo como un valor cercano al propio.

Dependiendo de cual sea la distribución de los datos, es muy posible que la media resulte ser un

Josu Mezo es profesor de la Universidad de Castilla-La Mancha y editor de *malaprensa.com* (editor@malaprensa.com).

dato que muy poca gente reconoce como corriente o mayoritario. Si calculamos la edad media de todas las personas que viven en una ciudad española (en Toledo, en 2002, era 37,7 años), todos comprendemos que, dada la distribución de edades en la población, la mayoría de la gente no tiene 38 años, ni cifras parecidas. Pero nadie se extraña, ni dice “esa media está mal”, porque el dato no se parezca al suyo o al de sus amigos y conocidos.

En cambio, si alguien afirma que el sueldo medio en España es de 20.200 euros brutos (lo era en 2007), no tardan en aparecer voces que dicen “pues yo no conozco a nadie que gane eso”, “pues alguien se debe de estar quedando con mi parte porque yo gano la mitad” y cosas similares, que encuentran eco a veces en los medios de comunicación. Pero es tan poco sorprendente que la mayoría de la población gane cantidades muy diferentes al sueldo medio, como que la mayoría de la población tenga edades muy diferentes a la edad media. Nadie debería extrañarse ni de lo uno ni de lo otro. El problema no es que la media diga una tontería, sino que hay gente que lee en ella lo que la media no dice.

Por otra parte, la estadística no se acaba con la media, como parecen creer algunos. La media es sólo una de las herramientas disponibles para describir la realidad. Y

Muy a menudo, sería más informativo acompañar o sustituir la media por otros indicadores.

aquí es donde está quizá el problema en la práctica cotidiana de nuestras instituciones productoras de estadísticas, y en el seguimiento periodístico de las noticias sobre ellas: que, muy a menudo, se utiliza únicamente la media, cuando sería más informativo acompañarla o sustituirla por otros indicadores.

La mejor alternativa sería seguramente la mediana, que es el valor que ocupa la posición central de una población, de forma que la mitad de los casos (restantes) tiene un valor inferior, y la otra mitad tiene un valor superior. Para averiguarla, sólo hay que ordenar los valores de todos los sujetos y tomar

el valor que ocupe la posición central. Así, el sueldo mediano de los empleados de una empresa lo podríamos averiguar poniendo en fila a todos los empleados, de menor a mayor sueldo, y preguntando cuánto gana al que ocupe la posición intermedia (en realidad, claro, no ponemos a nadie en fila, sino que pedimos al ordenador que nos diga cuál es el valor que ocupa esa posición central).

La mediana no resolvería el famoso ejemplo de los pollos, pero, en realidad, hay pocas situaciones sociales o económicas que tengan una distribución como esa, con todos los casos agrupados en dos valores extremos. Son mucho más comunes otro tipo de distribuciones. Por ejemplo, si representamos en un gráfico una variable física, como la altura de los españoles adultos, colocando en el eje horizontal las diferentes medidas en orden creciente, y sobre ellas sendas columnas, con altura proporcional al número de personas que tienen cada una de las tallas, nos saldría un dibujo parecido a una campana: muy poca gente muy bajita, un poco más de gente con cada centímetro que ascendemos, hasta llegar al valor más común (la parte alta de la campana) y, a partir de ahí, tendríamos menos y menos gente a medida que vamos subiendo centímetros. Sería el caso inverso al de los pollos (pocos valores extremos, muchos va-

lores medios), pero, de nuevo, la mediana no aportaría aquí gran cosa. De hecho, si la distribución fuera absolutamente simétrica, la media y la mediana tendrían el mismo valor.

Sin embargo, la mayoría de las variables sociales y económicas no tienen una distribución como las variables físicas, simétrica y en forma de campana, sino asimétrica. Imaginen un gráfico que representara el tamaño de las empresas españolas, por número de empleados. Habría muchísimas con uno, dos o tres empleados. Quizá el número más común sean dos o tres. Y, a partir de ahí, habría cada vez menos empresas con cuatro, cinco, diez, 20, 100, 500... El gráfico no sería una campana, sino más bien parecería un ratón, con el cuerpo en la parte izquierda y una larguísima cola hacia la derecha de valores cada vez más altos, con cada vez menor número de empresas que los tienen. Una gran cantidad de variables económicas darían lugar, al representarlas, a un gráfico en forma de ratón, con cola a la derecha: si pensamos en empresas, todos los que son indicadores aproximados de tamaño (trabajadores, ingresos, beneficios, impuestos pagados, etc.); casi todas las variables económicas sobre individuos y familias (salarios o ingresos, impuestos, gastos totales o gastos en diferentes partidas); los precios de muchos bienes o ser-

vicios (ordenadores, electrodomésticos, coches, viajes de vacaciones, bodas, seguros); o el tamaño o el precio de los pisos construidos, vendidos o alquilados, entre otros muchos ejemplos.

Pues bien, cuando tenemos una variable cuyo gráfico tiene forma de ratón, y no de campana, es cuando la mediana y la media se comportan de manera muy diferente. Veamos un ejemplo simplificado. Imaginemos una empresa con 50 trabajadores, con la siguiente distribución de salarios: 20 empleados ganan 800 euros al mes, otros 20 ganan 1.000, nueve ganan 1.500 euros y uno, el dueño, gana 30.000 euros mensuales. El sueldo medio es 1.590 euros. Una cifra en la que no se reconocerán la mayoría de los empleados (sólo nueve de los 50 ganan algo parecido), ya que la media se ve alejada de los valores comunes por los diez salarios más altos (la cola del ratón) y, en particular, por el sueldo del jefe, que hace la media poco “representativa”. En cambio, el sueldo mediano en esa empresa es de 1.000 euros (es el sueldo de los trabajadores que ocupan la posición 25 y 26, tras ponerlos en orden de menos a más sueldo). Esa sí es una cifra reconocible por la mayoría de los trabajadores como cercana al sueldo normal, ya que el sueldo mediano no se ve afectado, como el medio, por los valores altos de la cola.

En las distribuciones asimétricas, la mediana es un valor más cercano al de la mayoría de los sujetos que la media.

Generalizando, en las distribuciones asimétricas (más parecidas a un ratón con larga cola que a una campana), la mediana suele ser un valor más “representativo” que la media, en el sentido de que es un valor más cercano al de la mayoría de los sujetos. Esa diferencia no será siempre tan rotunda como en el ejemplo anterior, pero muchas veces es suficientemente relevante como para hacer preferible utilizar la mediana. Por eso, en muchos países, cuando se habla, por ejemplo, de los ingresos de las familias, se utilizan datos como el salario mediano o la renta familiar mediana.

Desgraciadamente, sin embargo,

en España, esta costumbre no está nada arraigada, a pesar de que, en muchos casos, las instituciones que generan y publican las estadísticas tienen datos individuales, y no sólo agregados, y, por tanto, representaría para ellas un esfuerzo ínfimo producir y difundir, junto con los datos medios, los datos medianos. Una excepción es la Encuesta de Estructura Salarial, realizada por el Instituto Nacional de Estadística (INE), que permite ver que, en 2006 (último año publicado), el sueldo bruto medio anual era en España de 19.700 euros, mientras que el sueldo mediano era de 15.700 euros. Es una distancia notable, que ilustra, con datos reales, el mismo fenómeno explicado más arriba para la ficticia empresa de 50 trabajadores. La mayoría de los trabajadores se sentirían más cercanos a la segunda cifra que a la primera si alguien les preguntara por un salario “representativo” (hay que tener cuidado, de todos modos, porque estas cifras mezclan salarios de trabajos a tiempo completo y tiempo parcial).

Otro concepto estadístico muy útil, relacionado con la mediana, es el del percentil, del que habrán oído hablar a pediatras y padres de niños pequeños. ¿Qué quiere decir eso de que el niño está en el percentil 35 de peso o en el percentil 73 de altura? Pues simplemente que el niño pesa más que el 35% de los

niños de su edad, o que mide más que el 73% de los niños de su edad. Se podría decir que la mediana no es nada más que un percentil especial (el percentil 50). Otros dos percentiles especiales son los llamados cuartiles, que son el percentil 25 y el percentil 75, es decir, aquellos valores que tienen por debajo el 25% y el 75% de los casos. En España, en 2006, los cuartiles de los salarios eran 11.900 y 23.300 euros. El INE nos da además los percentiles 10 y 90 para esta encuesta concreta, es decir, los valores que delimitan el 10% de los trabajadores con menos y con más ingresos. En 2006, fueron 8.200 y 34.900 euros (recuerden que se incluyen trabajadores a tiempo parcial).

Basta así, con la mediana y los dos cuartiles, para ver cuáles son los puntos de corte que dividen a la población en cuatro partes compuestas por el mismo número de personas, lo cual da mucha información sobre cómo se distribuye el asunto social y económico en cuestión. Estamos ya muy lejos de la broma de los dos pollos, y no hemos tenido que manejar ninguna herramienta estadística complicada. Y aunque es cierto que los lectores u oyentes pueden no saber qué es la mediana o los cuartiles, la explicación es tan sencilla que cabe en un pequeño paréntesis o nota al pie en un gráfico o tabla.

¿No creen que sería mucho más

informativo, en lugar de la media, disponer de la mediana (y los cuartiles) cuando hablamos sobre los precios de los pisos, las pensiones, los sueldos, los ingresos declarados a Hacienda, el tamaño de las empresas? ¿O cuando informamos sobre el número de alumnos por profesor en Secundaria, los años que tardan en terminar la carrera los universitarios, los días de espera para las operaciones, los enfermos por médico en el servicio de salud, el número de libros que la gente compra o lee cada año o el número de días que se toma de vacaciones fuera del hogar? En casi todos estos casos, las autoridades tienen las herramientas para calcular tanto la media como la mediana y los cuartiles y, sin embargo, suelen facilitarnos sólo la primera.

Propongo que intentemos luchar contra esa costumbre, empezando por algo tan simple como pedir a las fuentes que nos den ambos tipos de estadísticas. Obviamente, el INE, la Agencia Tributaria o el Instituto Nacional de Empleo (INEM) no van a cambiar de un día para otro sus rutinas. Son grandes instituciones con procedimientos estandarizados y repetitivos a los

Propongo que pidamos a las fuentes que nos den ambos tipos de estadísticas: media y mediana y cuartiles.

que les cuenta cambiar el curso. Pero si hay suficiente demanda, como el coste es mínimo, acabarán haciéndolo. Y otras instituciones más pequeñas (ayuntamientos, consejerías, empresas, instituciones de investigación) probablemente respondan a nuestras peticiones de manera mucho más flexible. Con un poco de suerte, en unos años, la broma de los dos pollos dejará de parecer realista. ♦