

Pueblos pequeños, números grandes

JOSU MEZO

A los vascos (¿y a quién no?) nos encanta leer en el periódico noticias que nos dicen que hacemos las cosas bien, o que somos los mejores en alguna cuestión. Una de las cosas en las que, por lo visto, el País Vasco destaca es la esperanza de vida, indicador esencial que puede considerarse una medida indirecta de muchos otros aspectos relacionados con el bienestar social. Y así, imagino que muchos vascos se sentirán complacidos al leer en sus medios titulares como estos, repetidos regularmente: “Las mujeres vascas son las más longevas de Europa, con 85,7 años” (*Elcorreo.com*) y “La esperanza de vida de las mujeres vascas es la más alta de Europa, con 86,2 años (*Elmundo.es*)”. El primer titular era de 2012 y el segundo de 2018, y por eso la diferencia de edades. Por lo demás, el relato permanece. Sin embargo, es falso, ya que las mujeres vascas no son ni siquiera las más longevas de España, como se puede comprobar en una rápida visita a la web del Instituto Nacional de Estadística (INE). Entre 2012 y 2017, la esperanza de vida femenina del País Vasco ha sido

todos los años menor que la de Navarra, Castilla y León, Madrid y, en ocasiones, que la de La Rioja y Aragón.

¿De dónde sale entonces la información publicada tan reiteradamente por los medios? Pues nada menos que del Instituto Vasco de Estadística (Eustat), del que quizá con inocencia los medios han creído que podían fiarse en estas materias. ¿Y cómo es que el Eustat difunde repetidamente una información tan errónea? Pues porque compara las estadísticas vascas con las de los Estados de la Unión Europea (UE). Y en efecto, los datos del País Vasco son ligeramente mejores que los del conjunto de España, que a su vez tiene los valores más altos de la UE. Por tanto, lógicamente, las estadísticas vascas son mejores también que las de cualquier otro país europeo. Pero no son, como hemos visto, mejores que las de otras comunidades autónomas españolas. Y no sabemos si serían también superadas por los de otras entidades subestatales similares, como las regiones de Italia y Francia y los estados federados alemanes.

Más allá de la posible intencionalidad simbólica de comparar los datos del País Vasco con los Estados de la UE y no con los de las regiones, podríamos decir que aquí el Eustat está haciendo trampas estadísticas, al comparar los datos vascos (referidos a unos dos millones de personas) con los de poblaciones en general mucho mayores, como las de los Estados de la UE. Precisamente, cuando se comparan datos de grupos pequeños con otros de grupos grandes, los valores más extremos y llamativos tienden a aparecer en los grupos pequeños. Y así, los récords, positivos y negativos, aparecen normalmente en los pequeños, por puro azar, por accidente; a pesar de lo cual, a menudo se interpretan como portadores de alguna información relevante.

Para entender esto, podemos imaginarnos que tomáramos una población grande, como una gran ciudad, y pudiéramos medir una variable básicamente aleatoria de todos los habitantes que van por la calle en un momento dado, como, por ejemplo, cuántas monedas llevan encima. Después, dividiríamos la población en grupos, creando un grupo por cada tramo de cada calle de la ciudad, definido por el cruce con otras calles. Habría miles de grupos, con números muy distintos de personas. En algunos tramos habría tal vez cinco o diez personas, en otros tramos habría 500 o 1.000 personas. Finalmente, calcularíamos el número medio de

monedas portadas por las personas de cada tramo de calle. Si la distribución de las monedas es realmente aleatoria, el número medio de monedas portadas por cada uno de los grupos grandes sería bastante similar, y similar también al número medio de monedas portadas por todas las personas de la ciudad. Es la “ley de los grandes números”. Las personas con muchas o pocas monedas se compensarían unas con otras y quedarían de todos modos diluidas entre la mayoría, que llevaría un número normal de monedas, de manera que todos los grupos grandes tendrían medias parecidas.

En cambio, los valores medios de los grupos pequeños serían mucho más variados. Por puro accidente, en algunos grupos formados por pocas personas caerían varias que llevaban muchas monedas, o muy pocas, y harían que el valor medio de ese grupo fuera especialmente alto o bajo. Es la “ley de los pequeños números”. De forma que, con seguridad, los récords de los tramos de calle por los que la gente lleva más y menos monedas en el bolsillo no corresponderían a concurridas calles comerciales del centro, sino, más bien, a tramos cortos de calles tranquilas de barrios residenciales en las que solo habría en ese momento tres, cinco, ocho personas. Esa misma lógica se aplica a cualquier estadística en la que comparemos medias o porcentajes de fenómenos correspondientes a colectivos

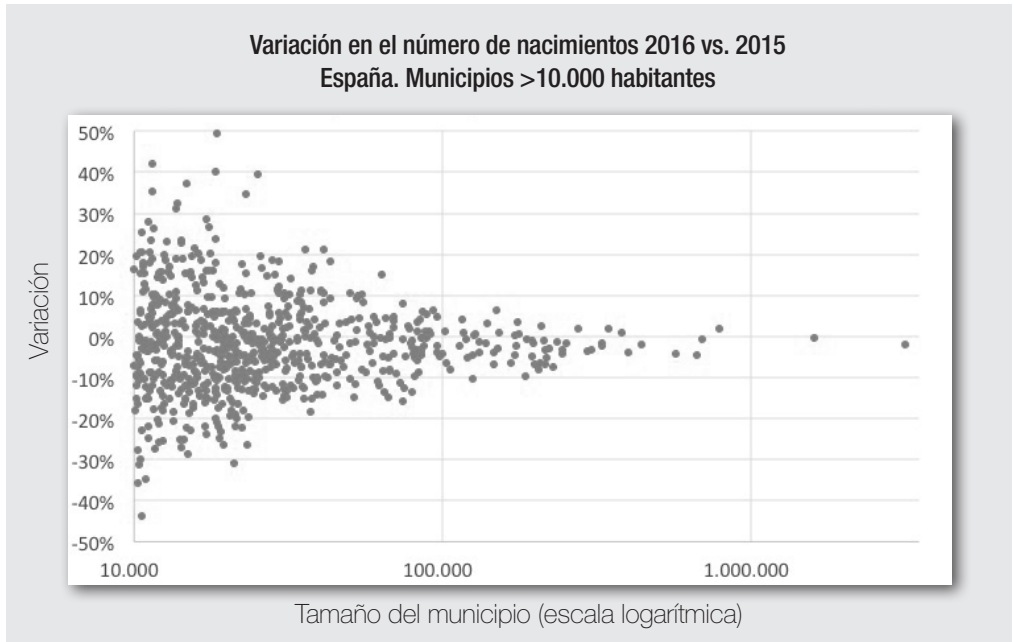
grandes y a colectivos pequeños. Estos últimos siempre tendrán los récords. Pero eso no significa (necesariamente) que esos colectivos tengan ninguna característica especial.

Veamos un ejemplo real. Con datos del INE es posible conocer el número de niños nacidos cada año en los municipios mayores de 10.000 habitantes. En el Gráfico 1 se representa cuánto varió ese número, en porcentaje, entre 2015 y 2016. Cuanto mayores son los municipios, más parecidos son sus datos, que son a su vez similares al valor prome-

dio del conjunto de los municipios (una caída del 2%). La media de las medias de los pueblos pequeños es también similar; no obstante, hay mucha mayor variedad, y por eso los valores más altos y más bajos están en municipios relativamente pequeños.

El gráfico en forma de cono tumbado es típico de muchas estadísticas en las que se comparan valores de colectivos pequeños y grandes. En él se ven en acción, por un lado, la ley de los grandes números, por la que los colectivos grandes tienden a parecerse entre sí y

Gráfico 1



Fuente: elaboración propia a partir de datos obtenidos de www.ine.es. Se usa como "lugar de nacimiento" el lugar de residencia de la madre

tener valores similares a la media total. Y también se ven los efectos de la ley de los pequeños números, por la que los colectivos pequeños, aunque en conjunto tienen una media parecida a la de los grandes, tienen valores más dispersos y, por tanto, suelen acaparar los récords tanto por arriba como por abajo.

Por todo esto, cualquier información del tipo “el pueblo de España donde sucede más (o menos) tal o cual fenómeno” debe ser tomada con mucha cautela. En algunos casos, obviamente, ese récord tiene que ver con verdaderas características de los municipios y responde a rasgos duraderos y estructurales, como el pueblo más rico, el más pobre, el que ha crecido más o menos en los últimos años y el que tiene una población más joven o más vieja. Pero, en muchos otros casos, las diferencias entre los pueblos son puramente accidentales, y los récords corresponderán de una manera aleatoria a pueblos pequeños.

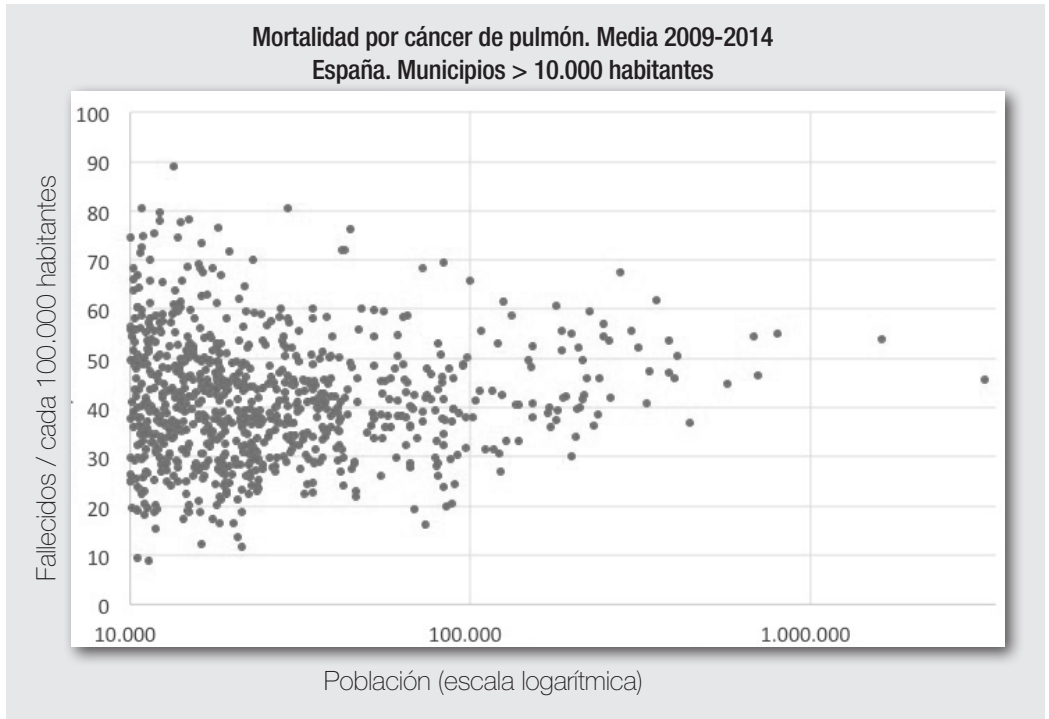
Así, por ejemplo, *El Confidencial* publicó el año pasado un reportaje sobre la alta incidencia del cáncer de pulmón en municipios industriales de Asturias, que está relacionada con la contaminación del aire causada por actividades industriales. La información se basaba, entre otras fuentes, en una base de datos facilitada por el INE sobre el número de fallecidos por cáncer de pulmón en los municipios españoles

entre 2009 y 2014 (por motivos de confidencialidad, el INE solo facilitó los datos de municipios mayores de 10.000 habitantes). Esto permitió al periódico acompañar el reportaje sobre Asturias con otra información separada sobre la incidencia del cáncer de pulmón en los municipios de toda España (mayores de 10.000 habitantes), en la que se podía buscar además la información sobre cualquier municipio para el año 2014. El antetítulo y el primer párrafo de esta pieza destacaban que Mora, en Toledo, fue ese año el pueblo español con la mayor tasa de fallecimientos por cáncer de pulmón. El texto también destacaba que Mula, en Murcia, fue el pueblo con menor incidencia (mortal) de esa enfermedad. Imagino que en ambos pueblos la noticia fue ampliamente difundida, en uno con preocupación y en otro con alegría.

Sin embargo, probablemente, no habría muchos motivos para la alarma ni para la satisfacción. *El Confidencial*, con transparencia ejemplar, puso a disposición de los lectores una base de datos con los resultados obtenidos del INE, y esto permite ver que en esta estadística también se cumplen las leyes de los pequeños y los grandes números. El Gráfico 2, con datos medios del periodo 2009-2014, lo muestra.

La gráfica tiene de nuevo la forma de un cono tumbado: los pueblos y ciudades grandes tienen valores similares

Gráfico 2



Fuente: elaboración propia a partir de datos del INE facilitados por *El Confidencial*

entre ellos, y próximos al valor central. Los pueblos pequeños tienen valores más dispersos. Ahora bien, que en un pueblo haya más o menos incidencia de cáncer de pulmón no es un hecho totalmente aleatorio¹. Y por eso, proba-

blemente, se observa mayor dispersión que en el gráfico anterior para todos los tamaños de municipios. En particular, en la parte superior se deja ver una estela de municipios, de diversos tamaños, bastante alejados de los demás,

¹ En realidad, como ya señalaba *El Confidencial*, lo correcto sería utilizar tasas de mortalidad ajustadas por la edad, las cuales eliminan el efecto que tienen sobre las tasas crudas las diferentes estructuras de edad de los municipios (dos municipios con el mismo riesgo para cada edad tendrán tasas crudas diferentes si tienen estructuras de edad diferentes). Tal vez algunos de estos municipios salgan perjudicados en la tasa cruda por estar relativamente envejecidos. Sin embargo, las tasas ajustadas por edad no están disponibles en el ámbito municipal, por lo que no es posible hacer esa comprobación

que parecen tener valores anormalmente altos. Entre ellos están algunos pueblos y ciudades donde efectivamente podemos sospechar que hay factores reales que explican su alta posición, por su presente o pasado industrial, como Sestao, Langreo, Basauri, Mieres, Avilés, Ferrol, Barakaldo y Gijón.

No obstante, en la banda alta como sobre todo en la banda baja predominan los pueblos pequeños. La mayoría de ellos están allí probablemente por accidente. Y de hecho, en ese promedio de seis años, ni Mora ni Mula ocupan posiciones destacadas. Sus valores extremos de 2014 eran, como cabía esperar, puros accidentes derivados de la ley de los pequeños números. Están entre los pueblos pequeños de esta base de datos, y por eso, por puro azar, en un año concreto tuvieron datos atípicamente altos o bajos, los cuales no significan nada importante ni transmiten información verdadera acerca de que pase algo especial, bueno o malo, en esos municipios.

La misma lógica se puede aplicar a todo tipo de noticias en las que se comparan valores de colectivos con tamaños muy distintos: los datos de un pueblo con medias regionales o nacionales, los de un conjunto de empresas pequeñas con otro de empresas grandes y los resultados de hospitales de

pequeñas ciudades de provincias con los de las grandes capitales. Los valores mejores, en aquella cosa que estamos midiendo, estarán casi siempre entre los pueblos, empresas y hospitales pequeños. Pero, atención, también estarán entre los pequeños los valores peores.

Por ello, haciendo una trampa similar a la de la noticia sobre el País Vasco con la que empezaba este texto, tiene poco valor la información publicada en mayo de este año en la que 20 colegios privados españoles presumían de que sus alumnos habían obtenido mejores resultados que los países que lideran el informe PISA. Claro, un pequeño grupo de colegios privados, seleccionados entre los de élite en España, tiene mejores resultados que la media de países enteros. ¡Lo raro sería que no fuera así!

Precisamente, en torno al tema de los centros educativos, se produjo hace unos años en Estados Unidos un famoso episodio de mala interpretación de los datos, con consecuencias esta vez nada triviales, por ignorar la ley de los números pequeños. Al parecer, varios estudios sobre el rendimiento de los centros escolares observaron que los más pequeños estaban sobrerrepresentados entre los colegios o escuelas con mejores resultados en los test estandarizados². Estos resultados

2 Nótese que se habla de tamaño de los centros, no de los grupos o las aulas, asunto muy distinto

se repetían en diferentes contextos, lo que llevó a la Fundación Bill Gates y a otras instituciones a gastar cientos de millones de dólares en programas para mejorar el sistema educativo, que incluían la división de centros escolares grandes en varios más pequeños³. Sin embargo, todo se basaba en una mala comprensión de los datos, ya que los centros más pequeños también estaban sobrerrepresentados entre los que tenían peores resultados escolares. Es decir, cumpliendo la ley de los pequeños números, había muchos centros pequeños entre los que tenían notas anormalmente altas y bajas. Ser

un centro pequeño no te hace mejor. Solo hace más probable que, de vez en cuando, tus estadísticas sean “raras”. Tras descubrirse el (caro) error, las ayudas a la fragmentación de centros se fueron eliminando. En cambio, la moraleja permanece: las escuelas pequeñas, como los pueblos pequeños y las empresas pequeñas, tienden a producir números grandes (positivos o negativos), con los cuales es necesario ser cuidadoso al interpretarlos. Si no lo hacemos, podemos perder, tal vez no millones de dólares, pero sí nuestra buena comprensión de la realidad.

3 Véanse más detalles en: Tabarrok, Alex. 2010. “The Small Schools Myth”. *Marginalrevolution.com*, 2 de septiembre de 2010 (<https://marginalrevolution.com/marginalrevolution/2010/09/the-small-schools-myth.html>). Visitado el 6 de noviembre de 2018