

Espejismos estadísticos

JOSU MEZO

Hace unos años, en una visita a Málaga, encontré en un periódico local un ejemplo maravilloso de una paradoja que, a pesar de ser conocida por los estadísticos y fácil de explicar, puede llevar a errores importantes en la interpretación de datos. Estábamos todavía en plena burbuja inmobiliaria (principios de 2006), pero el periódico informaba de que en 2005 el precio medio por metro cuadrado de las viviendas turísticas en la provincia de Málaga había bajado un 13,6% respecto a 2004. Era, en aquel entonces, una noticia muy extraña. Los precios seguían subiendo en el conjunto de España, tal vez un poco más despacio que en años anteriores, si bien aún a un ritmo considerable. ¿Qué estaba pasando en Málaga? ¿Tal vez estábamos ante el comienzo del fin del *boom*, que comenzaba a manifestarse en las viviendas turísticas?

Pues no, nada de eso ocurría. De hecho, la propia noticia contaba que el precio por metro cuadrado seguía subiendo en la mayoría de los municipios de las zonas turísticas de la provincia. ¿Cómo era posible entonces

que el precio medio de la provincia hubiera bajado? La razón era que de un año a otro había cambiado de manera importante el número de pisos disponibles en distintas zonas; y, en particular, había disminuido mucho la cantidad de viviendas en oferta en Marbella, que era precisamente, con gran diferencia, el municipio más caro de la provincia. De manera que, aunque, municipio a municipio, los precios subían, la media provincial de 2005 era más baja que la media de 2004, al entrar en la media de 2005 muchos menos pisos del municipio más caro.

Veamos otro ejemplo similar sobre un tema muy diferente. En Estados Unidos, como en todos los países desarrollados, llevan años preguntándose por el rendimiento de su sistema educativo. Ya en los primeros años 70, empezaron a hacer pruebas de evaluación en miles de centros educativos (National Assessment of Educational Progress –NAEP–) para medir, a distintas edades, los conocimientos de los alumnos en diferentes materias. Si uno mira la evolución de las notas globales, el

resultado es bastante descorazonador. Los valores medios apenas han mejorado nada en 40 años. Por ejemplo, a los 17 años, no ha habido una mejora estadísticamente significativa en las notas medias de lectura y de matemáticas. Sin embargo, cuando se analizan los datos por grupos raciales, la historia es distinta: tanto en matemáticas como en lectura, todos los grupos han mejorado. Especialmente, han mejorado los grupos más desfavorecidos, los negros y los hispanos, cuyas notas han subido aproximadamente un 10% y han reducido su distancia con las de los alumnos blancos.

Pero, si sube la nota de cada grupo de población, ¿cómo no sube la media del conjunto de los grupos? Pues por el cambio en la composición de la población estudiantil. En los años 70, los blancos, el grupo con mejores notas, eran más del 80% de los estudiantes de 17 años. En 2012, eran solo el 56%. Por el contrario, la proporción de estudiantes negros ha subido ligeramente y los hispanos, que eran menos del 5% a primeros de los 70, han pasado a ser el 22% en 2012. Y así, aunque las notas de cada grupo mejoran, al haber ahora más alumnos de los grupos con notas relativamente peores, la nota media global no ha mejorado.

Los dos casos son muy similares y tienen que ver con lo que a veces se

llama efecto composición o sesgo de composición. Este se produce cuando comparamos la evolución a través del tiempo de un indicador, como la media, referido a un cierto grupo de casos (pisos, estudiantes, empresas, trabajadores...), cuya composición ha cambiado alguna característica que está relacionada con el valor que estamos midiendo. Es decir, dentro del grupo general, hay subgrupos con valores distintos, y esos subgrupos han cambiado de tamaño relativo. En el caso de los pisos, la característica que variaba era la distribución de su localización. En el caso de los estudiantes, lo que se modificaba era la composición racial (y, por asociación, socioeconómica). El efecto composición produce lo que podríamos llamar un espejismo estadístico: parece que los precios bajan, cuando, en realidad, están subiendo; parece que los conocimientos no suben, cuando lo están haciendo en todos los grupos.

El efecto composición está también seguramente detrás de muchas estadísticas chocantes en relación con la evolución de los salarios. No era raro ver noticias en 2006 o 2007 en las que se comentaba, con extrañeza, que los salarios en España apenas subían, o incluso bajaban, a pesar de una década de bonanza, crecimiento económico y paro en descenso. Por el contrario, también se publicaron noticias en los primeros años de la crisis en las que se

aseguraba que, pese a la fuerte caída del paro, los salarios medios estaban subiendo en España. En situaciones así uno siempre tiene que pensar en la posibilidad de que estemos ante una manifestación del efecto composición. Es decir, en este caso, tendríamos que preguntarnos cómo ha podido cambiar en estos años la composición del empleo en España y la composición de la población trabajadora.

Y así, en los años de bonanza, probablemente se crearon muchos empleos poco cualificados, con sueldos relativamente bajos. De manera que, aunque los sueldos de cada categoría laboral subieran durante aquellos años, los sueldos medios podrían estar bajando. Esto es, por simplificar, si tanto los camareros como los ingenieros ven subir sus sueldos, pero el número de camareros aumenta mucho más que el número de ingenieros, es posible que el salario medio baje. Podemos resumir esa situación como “los sueldos están bajando”, si bien no es una afirmación correcta del todo. Más bien, deberíamos decir algo como “cada vez hay más empleos de los de sueldos bajos”. Asimismo, en aquellos años, entraron en el mercado de trabajo muchos inmigrantes y mujeres, las cuales se dedicaban en periodos anteriores a las tareas domésticas. Al menos, en cuanto a su experiencia laboral, y probablemente por otros factores, como la formación,

estos nuevos trabajadores tendrían características que los llevarían a ganar normalmente salarios más bajos. Así pues, podríamos estar también ante un efecto composición relacionado con las características de la población ocupada.

Del mismo modo, seguramente, el efecto composición explicaba la aparente “subida de los sueldos” de los primeros años de la crisis. Las empresas eliminaron más empleo de la gente más joven, con menos experiencia, con contratos temporales o con menor indemnización. Por lo tanto, en general, personas con sueldos medios más bajos. Y así, aunque la mayoría de las personas que siguieron trabajando vieran sus sueldos congelados, o incluso reducidos, la media de los sueldos de los que seguían trabajando pudo ir subiendo, respecto a la media de todos los que trabajaban unos años antes. En este caso, en el que el grupo se transforma porque algunos de sus miembros desaparecen, estamos ante una variante especial del sesgo de composición que se llama sesgo de supervivencia.

El efecto composición no es relevante solo para comparaciones a lo largo del tiempo. Igualmente, puede manifestarse cuando comparamos, en un mismo momento, los datos de dos o más entidades o grupos que difieren en su composición interna en alguna característica relevante para el tema estudiado.

Volviendo al tema de los salarios, es habitual ver estadísticas sobre el salario medio en diferentes comunidades autónomas que nos informan, por ejemplo, de que los sueldos medios en Canarias son un 30% más bajos que en Madrid. Quizá, en ese caso, la mayor parte de los lectores sea consciente de ello, pero no está de más recordarlo: esa diferencia se explica en parte porque por “el mismo trabajo” (cualificación del trabajador, experiencia, jornada, sector, tamaño de la empresa...) se cobra menos en Canarias que en Madrid; no obstante, también se justifica en buena parte porque los trabajos que hay en Madrid son distintos de los de Canarias. Parte de la explicación es que el camarero y el ingeniero que trabajan en Canarias ganan menos que los que trabajan en Madrid, y también que, proporcionalmente, en Madrid hay más ingenieros y menos camareros que en Canarias.

Otro campo en el que el efecto composición puede llevar a conclusiones erróneas al lector ingenuo de las estadísticas es el de las investigaciones sobre resultados educativos en diferentes tipos de centros o con diferentes métodos de enseñanza. Hay mucha evidencia de que los resultados educativos de los niños y jóvenes están muy asociados a características personales y familiares, que son previas e independientes de lo que pase en la escuela.

Entre ellas, están el estatus económico, que va asociado a condiciones materiales de vida, el nivel educativo de los padres o el conocimiento del idioma de enseñanza (importante en el caso de niños y jóvenes inmigrantes). Naturalmente, si comparamos los resultados académicos de centros educativos o de tipos de centros (privados o concertados versus públicos, bilingües o trilingües versus monolingües, religiosos versus aconfesionales), sin tener en cuenta la composición de su alumnado (por clase social, por estudios de los padres, por lengua materna), obtendremos resultados de poco valor y sesgados positivamente hacia los centros con alumnos de clase social más alta, con padres con mayor nivel de estudios y con menos población con problemas de inserción por motivos lingüísticos.

Ello no quiere decir, por supuesto, que al tener en cuenta el efecto composición desaparecerán todas las diferencias entre centros educativos. Seguirá habiendo probablemente unos centros que tienen resultados mejores que otros, a igualdad de características de su alumnado. Lo que importa es darse cuenta de que solo entonces, tras eliminar los sesgos de composición, estaremos ante información realmente valiosa acerca de las formas de gestión, métodos de enseñanza u otras características relevantes de los centros sobre las que nos podría interesar actuar para

mejorar los resultados de la enseñanza.

Podrían ponerse ejemplos en muchos otros ámbitos de informaciones que comparan “en bruto” datos de dos grupos y llegan a conclusiones apresuradas, sin preguntarse si las diferencias pueden explicarse por el efecto composición. Por ejemplo, se comparan a veces datos de accidentalidad de jóvenes y viejos, u hombres y mujeres, sin hacerse las preguntas adecuadas sobre las diferencias en el comportamiento de unos y otros. Así, las mujeres conductoras tienen muchos menos accidentes que los hombres, y todos tenemos evidencia anecdótica que nos hace pensar que los conductores más imprudentes suelen ser varones. Sin embargo, también es probable que, en promedio, las mujeres conduzcan menos kilómetros al año que los hombres. Una buena comparación debería tener esto en cuenta, y confrontar el promedio para hombres y mujeres de accidentes/100.000 km conducidos, o algo similar. Seguramente, aún así, las mujeres resultarían conductoras más seguras, pero lo cierto es que sin ese tipo de dato no podemos afirmarlo.

Según otra información, hace unos años, la Dirección General de Tráfico estaba especulando con la idea de prohibir la conducción a conductores noveles (carné de menos de un año) entre las 23:00 y las 06:00 horas,

porque estos tenían el 18% de sus accidentes durante esas horas, mientras que los demás conductores solo tenían el 12% de sus accidentes en ese tramo horario. Más allá de si la diferencia es lo suficientemente grande como para justificar una medida tan estricta, el problema principal de esa estadística es que no parece tener en cuenta la posibilidad de que los conductores noveles (normalmente, jóvenes) sean mucho más frecuentemente conductores ocasionales, de ocio, especialmente de ocio nocturno, que los demás conductores. Por lo tanto, de nuevo, para hacer conclusiones razonables, habría que comparar la accidentalidad de los noveles y el resto en relación con las horas conducidas dentro y fuera de ese horario nocturno.

En fin, los ejemplos podrían seguir indefinidamente. El mensaje común sería que cuando veamos que diversas personas (o empresas, escuelas, pisos...) agrupadas por una cierta característica A tienen un valor medio diferente en otra característica B, debemos tener cuidado para no concluir, precipitadamente, que hay una relación directa entre A y B. En particular, debemos preguntarnos si puede ser que haya otras características en las que los grupos sean diferentes y si, al tenerlas en cuenta, la asociación entre A y B se mantiene o desaparece. Podríamos decir que, en el fondo, estamos otra vez

ante el viejo consejo de comparar peras con peras y manzanas con manzanas. Por continuar la metáfora, se trataría de preguntarnos si la diferencia de precio entre unas manzanas golden y otras reineta se explica solamente por

la variedad de la manzana, o también puede ser que haya entre ellas otras diferencias: por origen, por brillo, por apariencia... Si no nos lo preguntamos, podemos ser víctimas de un evitable espejismo estadístico.