

La Voz de Galicia, 15 de Mayo de 2000

2000 Año Mundial de las Matemáticas

La voz de la Escuela

Ricardo Cao Abad Antes de las vacaciones hablábamos de la estadística y algunas relaciones con la probabilidad. Otro de los puntos en los que la probabilidad ofrece una ayuda importantísima a la estadística es el muestreo, que se define como la parte de la estadística que se dedica a diseñar mecanismos aleatorios para recoger observaciones con el fin de obtener un valor aproximado de determinada cantidad de interés.

Elegir una muestra

Uno de los principios básicos de elección de una muestra es utilizar un mecanismo aleatorio (observar una ruleta, lanzar un dado, o usar un generador de números aleatorios de un ordenador, por ejemplo) mediante el cual todas las posibles muestras tengan la misma probabilidad de salir elegidas. Esto es muy fácil de decir pero a veces muy difícil de conseguir. De hecho este es uno de los problemas que surge muy a menudo a la hora de realizar encuestas de opinión o sondeos electorales. Las cosas pueden ser especialmente peligrosas cuando no tenemos ningún control sobre el método de elección de la muestra. Veamos un ejemplo.

Los vecinos del primero

Día tras día vengo observando cuando bajo en ascensor desde mi casa (en un octavo piso que es el último del edificio) como, con cierta frecuencia, ocurre que mi ascensor se para en otro piso y nadie se sube. Sin duda obedece a la desconsideración de otros vecinos que llaman los dos ascensores del edificio y, al llegar el otro antes, lo cogen. Además estoy casi seguro de que esto me sucede con mucha mayor frecuencia en los pisos más bajos. Parece como si mis vecinos del primero fuesen más desconsiderados que los demás. Para comprobarlo, decido anotar durante unas semanas el número de veces que el ascensor se para en cada piso y desglosar los resultados según se suba alguien o no.

Examinando los porcentajes de veces en que el ascensor se para y nadie se sube a él parece que se corrobora mi teoría, siendo los más desconsiderados los vecinos de los tres primeros pisos. Realmente esto parece muy raro. ¿No podría ayudarnos la probabilidad a entenderlo mejor?

Si pensamos un poco, vemos que los vecinos solidarios siempre aparentan lo que son. Como nunca llaman los dos ascensores, nunca ocurrirá que mi ascensor se pare en su piso (habiéndolo llamado ellos) sin que ellos se suban. Por el contrario, un vecino desconsiderado puede pasar por buen vecino, ya que es posible que, aún llamando los dos ascensores, llegue el mío antes a su piso, sin darme yo cuenta de ello. Pero, ¿cómo cuantificar todo esto?

CONTROL DE TURISTAS

Situaciones como las descritas en esta página ocurren muy a menudo en la vida real. Son los llamados datos sesgados por longitud o por tamaño. Por ejemplo, algo parecido sucede si tratamos de estimar el tiempo medio de estancia de un turista en nuestro país encuestando a los que están alojados en un hotel. Este mecanismo produce sesgo, pues es más probable elegir a un turista que pasará muchos días en el hotel que a otro que estará pocos. Para solucionarlo podríamos tener en cuenta el sesgo y corregir la estimación de la media o bien

realizar un muestreo no sesgado. Por ejemplo elegir aleatoriamente a los turistas según van pasando por la aduana. Así, todos tienen la misma probabilidad de ser elegidos.

UN EDIFICIO MUY UNIFORME

4 personas 300 kilos Supongamos, para simplificar, un edificio en el que en cada piso habitan sólo dos vecinos, uno considerado (nunca llama a ambos ascensores, sino sólo al mío -que es el que no va al garaje- y espera a que llegue) y otro no. De esta forma, la probabilidad de que un vecino de cierto piso (da igual de cuál) sea desconsiderado es de 0.5 (un 50%).

Así, si un vecino desconsiderado del séptimo llama ambos ascensores justo en el momento en que yo, desde el octavo, comienzo a bajar, con toda seguridad llegará antes mi ascensor que el otro, o, como mucho, a la vez y quedará, a mis ojos, como un buen vecino.

Si el vecino desconsiderado es del sexto, sólo observaremos su falta de solidaridad si el otro ascensor estaba en el séptimo o en el quinto. Para los otros seis pisos (bajo, primero, segundo, tercero, cuarto y octavo) el nuestro llegará al sexto antes, o, quizá, a la vez, que el otro ascensor. Se tiene, por tanto, que la probabilidad de que nuestro ascensor se pare en el sexto sin que nadie se suba es de $2/8$. En términos porcentuales, un 25%.

Razonando de la misma forma podemos calcular esas probabilidades para cualquier otro piso. De todas formas, como sólo uno de los dos vecinos de cada piso es desconsiderado, cuando pensamos que el que llama el ascensor es uno de los dos vecinos (aleatoriamente elegido) que habitan en ese piso, esas probabilidades se reducen a la mitad.

A la vista de los valores de la tabla resulta totalmente razonable que los datos que he recogido provengan de un edificio «uniforme», en el que, en cualquier piso, la probabilidad de ser considerado es la misma. ¡Ahora nuestros datos son indicativos de que la falta de solidaridad reina por igual (en torno a un 50%) en todos los pisos del edificio!

La pregunta clave es ¿cómo evitar sistemáticamente nuestra aventurada conclusión inicial? Hay dos formas de proceder. Una consiste en modificar el mecanismo de muestreo para obtener muestras no sesgadas de los porcentajes de vecinos insolidarios en cada piso. Esto puedo hacerlo situándome en cada piso y observando cuántos de los vecinos que van al ascensor adoptan una u otra conducta. De todas formas el hecho de ponerme a observar posiblemente modificará la conducta de los desconsiderados. El otro método consiste en corregir las estimaciones construidas a partir de estos datos sesgados, teniendo en cuenta el sesgo que tienen. Sabemos que sólo observamos algunos de los casos de vecinos desconsiderados y sin embargo todos los considerados los vemos siempre. De esta manera, la probabilidad de que si el ascensor se para en el primero, no se suba nadie es sólo el 87.5% de la probabilidad de que un vecino del primero sea desconsiderado. Es decir observamos como tales solo el 87.5% de los desconsiderados y ninguno de los considerados. Esto provoca que estimemos la probabilidad de interés como la frecuencia observada de veces en las que nadie se sube en el primero, dividida entre 0.875. Algo parecido ocurre para los demás pisos.