



La estadística, el

Observa esta imagen parcialmente cubierta y trata de adivinar qué es.

Cuesta trabajo, ¿no? La proporción del área de la foto que está descubierta es relativamente grande, pero no ayuda mucho. Puede haber varias razones. Una es que lo que se revela no sea tan importante; otra es que no estamos familiarizados con el total del paisaje. ¿Qué proporción de la foto tendría que estar descubierta para poder decir qué es con cierto grado de confianza?

Ahora mira la siguiente foto, que es la misma que la anterior, pero con la diferencia de que se han descubierto otras partes al azar. Trata de adivinar qué es nuevamente. Observa que es la misma proporción de área descubierta que en la primera foto, es decir, cuatro rectángulos de un total de 25.

Esta vez fue mucho más fácil proponer una solución con cierta confianza a pesar de que sigue cubierto más del 80% de la imagen. ¿Por qué? Ambas fotos son iguales y tienen descubierta el mismo número de rectángulos. Sin embargo, no son comparables en cuanto a la facilidad con la que la podemos figurarnos la foto completa, la

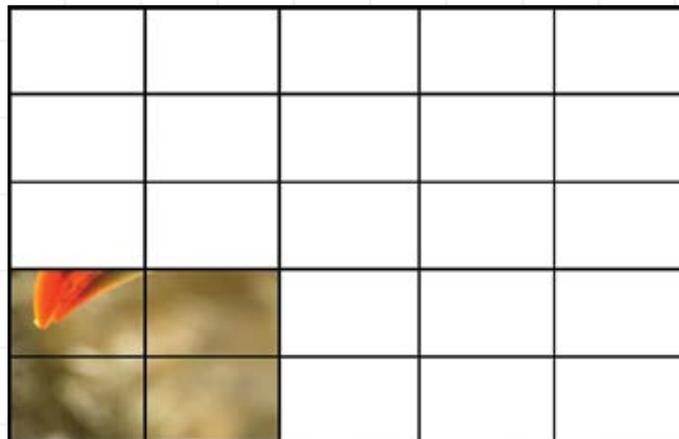


Figura 1.

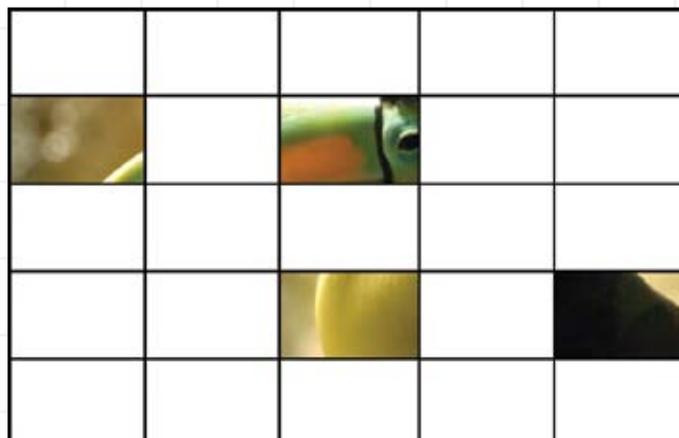


Figura 2.

cual se puede ver al final de este artículo. En el segundo caso, la parte visible es mucho más informativa que en el primero.

En la primera foto los cuatro rectángulos se eligieron contiguos y de manera que

estuvieran en la esquina inferior izquierda. En contraste, en la segunda foto los rectángulos fueron seleccionados al azar. Imagínate que se numeran los rectángulos del 1 al 25. En una bolsa de papel se meten 25 papelitos numerados, se revuelven bien y se sacan cuatro que determinarán qué rectángulos quedan a la vista. Hay muchas maneras de que salgan cuatro rectángulos del total y la segunda foto muestra una de ellas. Es relativamente fácil calcular que el total de maneras posibles de seleccionar cuatro rectángulos entre 25 es 12650. Todas ellas tienen la misma probabilidad de salir cuando uno extrae cuatro papelitos de la bolsa. Muchas serán informativas y pocas no lo serán, como nuestra primera foto. Así, es mucho más probable obtener un conjunto de rectángulos que sí sea informativo sobre el total cuando escogemos al azar qué rectángulos dejar visibles. A veces el caos es mejor que el orden. Sorpresas del azar.

Mientras más, mejor

Si tomamos una foto aérea, no podemos elegir dónde estarán las nubes y qué parte de la información cubrirán. Por eso es de

azar y otras sorpresas

Ignacio Barradas y Eloísa Díaz-Francés

interés saber cómo obtener información parcial para maximizar lo que nos revelará sobre el total. El ejemplo del tucán muestra que, cuando no se tiene acceso a toda la información de un fenómeno, en general es mejor elegir partes al azar que usar cualquier otro método. Éste es el tipo de problemas que la estadística nos ayuda a resolver. Por ejemplo, uno puede querer saber cuánta gente compraría un producto o votaría por un candidato. Pero en general es imposible preguntarle a toda la población. Lo que se hace es preguntarles a unos cuantos, y la mejor forma de escoger a quiénes es elegir al azar entre la población pertinente; por ejemplo, los mayores de 18 años en el caso de una campaña política.

La respuesta usualmente se expresa en términos de un intervalo de valores (entre 50 y 60% de los votantes apoyarían a Fulano) que se escoge de tal manera que sea alta la probabilidad de que incluya al porcentaje real. O dicho de otra manera, se cuantifica no sólo la probabilidad de que algo ocurra, sino también la incertidumbre o el error probable de la predicción: de alrededor del 90% de los votantes, más o menos un 3% votarían por Fulano.

Las probabilidades se expresan como valores entre 0 y 1 (o como porcentajes): lo que tiene probabilidad cercana a uno ocurre con mucha frecuencia y lo que tiene probabilidad cercana a cero no ocurre casi nunca. Volviendo a la foto del tucán,

si muchas personas trataran de adivinar qué es, no es difícil ver que, con pocos rectángulos descubiertos, serán pocas personas las que le atinen; mientras más rectángulos estén descubiertos, será mayor el número de personas que adivinarán. En el primer caso, se adivinará con poca frecuencia (probabilidad cercana a 0) y en el segundo con mucha frecuencia (probabilidad cercana a 1 o porcentaje cercano a 100%). En términos técnicos se dice que la incertidumbre disminuye con la información.

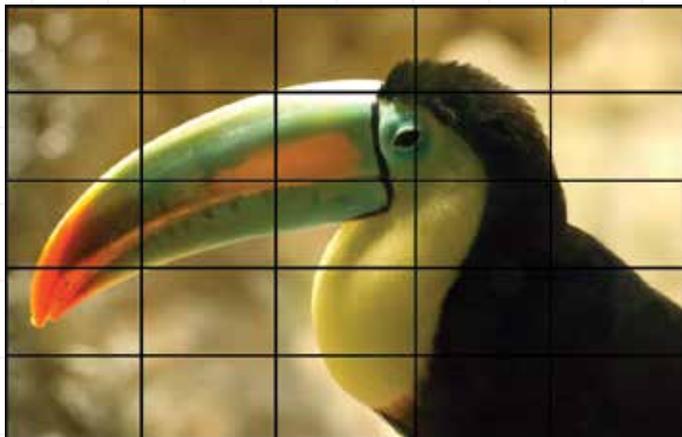
Por ejemplo, ¿qué color de coche es el más usual en tu ciudad? Para saberlo, podrías salir a la calle y anotar el color de todos los coches que ves en tu barrio. El problema con esta técnica es que tu ubicación no es al azar. Quizá vives cerca de un sitio de taxis y ves más coches de color taxi. Para hacerlo bien tendrías que idear algún método aleatorio para escoger un cruce de calles desde donde observar. Con un número total de coches observados y un número para cada color, podrías usar la estadística para proponer un intervalo de estimación para la verdadera proporción de coches de ese color que hay en la ciudad. Por ejemplo, podrías concluir que la proporción de coches rojos está en el intervalo (22-26%) con una confianza del 95%. Otra manera de expresarlo sería decir que hay una proporción estimada del 24% de coches rojos, más o menos un 2%. Así, el ancho de tu intervalo de estimación

es de 4%, que es también la precisión o certeza de tu afirmación acerca de los coches rojos. Decir que tu intervalo tiene una confianza o probabilidad de acertar de 0.95, o del 95%, equivale a decir que, si hubieras repetido la observación muchas veces, en aproximadamente el 95% de las veces el intervalo obtenido incluiría la verdadera proporción de coches rojos en la ciudad y que solamente alrededor de 5% de las veces no la contendría.

Entre más coches cuentes (entre mayor sea tu muestra), más angosto será el intervalo de incertidumbre, pero la confianza seguirá siendo igual. O equivalentemente, con una muestra mayor, un mismo intervalo de estimación tendrá más probabilidades de incluir el valor verdadero. Por eso vale la pena aumentar el esfuerzo y observar muchos coches para mejorar tus estimaciones.

La importancia de una s

Antes de continuar es importante resaltar una diferencia entre conceptos que en ocasiones se utilizan como si fueran intercambiables: estadística y estadísticas. Hay palabras que en plural no significan lo mismo que en singular. Aunque esto suene raro, una cosa es la estadística y otra muy diferente las estadísticas de algo. Las estadísticas son una colección de datos de cualquier tipo: cifras sobre tamaños de poblaciones, producción de diversos productos, resultados deportivos. Por ejemplo,



de acuerdo con el Instituto Nacional de Estadística y Geografía (INEGI), la población de México en 2010 era de 112 337 000 habitantes, con una tasa de crecimiento del 1.8% anual, lo que significa que se duplicará en 40 años. También nos dice que hay 95 hombres por cada 100 mujeres, que el 77.8% de la población es urbana y el número promedio de habitantes por casa es de 3.9. Estos son algunos de los muchos datos que se pueden encontrar en la página www.inegi.org.mx, en el rubro de población. En Internet puedes encontrar estadísticas acerca de prácticamente cualquier tema que se te ocurra.

En contraste, la estadística es una rama de las matemáticas que se dedica a reunir información y encontrar patrones en los datos que se obtienen de fenómenos aleatorios repetibles, cuantificando la incertidumbre en términos de probabilidad. La estadística nos sirve para entender mejor el mundo que nos rodea y poderlo predecir, si bien con incertidumbre. En ocasiones se buscará poder controlar o cambiar estos fenómenos, una vez que se conozca mejor su comportamiento aleatorio.

Por ejemplo, al estudiar con estadística los niveles de agua de un río a lo largo del año midiéndolos cada semana es posible describir muy bien, con alta probabilidad, los valores máximos de niveles de agua que se podrían alcanzar durante un acontecimiento extremo como un huracán. Lo sorprendente es que se puede decir mucho de estos valores extremos sin siquiera haberlos observado. Basta con estudiar los

los niveles habituales de agua. A partir de este comportamiento normal, la estadística permite cuantificar la probabilidad de que alcance otros niveles, e incluso se puede estimar aproximadamente cada cuántos años van a ocurrir estos acontecimientos extremos. Tomando esto en cuenta, los ingenieros civiles pueden calcular la altura ideal de un dique: suficientemente alto para contener el río en caso de que alcance niveles extremos, pero no tanto que sea demasiado caro.

Fiesta estadística

Se ha dicho que la estadística es el lenguaje de las ciencias naturales. Grandes avances científicos y también tecnológicos se han logrado gracias a esta disciplina. Es fundamental para conservar regiones naturales y proteger especies en peligro de extinción: las poblaciones de distintas especies se pueden estimar por métodos estadísticos sin necesidad de localizar a cada individuo de la especie. Sólo hay que escoger al azar unos cuantos puntos de observación dentro del

valores usuales del agua del río a lo largo del año, que son mucho menores en magnitud, pero que contienen mucha información sobre los posibles valores máximos. Es decir, no se necesita esperar a que llegue un huracán, basta con entender bien el comportamiento aleatorio del río con

hábitat de la especie y contar. Esta información sirve para tomar las decisiones pertinentes para la conservación de la especie.

Este año fue declarado Año Internacional de la Estadística. El objetivo de esta celebración mundial es aumentar la conciencia general sobre el impacto y utilidad de la estadística en todos los aspectos de nuestra sociedad. Hay muchas actividades interesantes, así como videos, animaciones, conciertos de música aleatoria y conferencias de divulgación. Puedes encontrar información en la página www.estadistica2013cimatax.mx.

La marca del azar

La estadística estudia fenómenos aleatorios que están regidos por el azar. Sin embargo, veremos con la siguiente anécdota que no es tan fácil entender cómo se comporta el azar, porque a veces éste resulta sorprendente.

Alex y su amigo Jorge decidieron echar 100 volados cada uno. Alex se puso un buen rato a lanzar una moneda. La lanzó 100 veces y registró lo que salía (A para águila y S para sol). En cambio, a Jorge le dio flojera y prefirió inventar los resultados para acabar más pronto. Aquí mostramos dos tablas, una con los resultados verdaderos de Alex y otra con los resultados falsos (imitación del azar)

A	A	S	A	A	A	S	S	A	A
S	S	S	S	A	A	S	A	S	A
S	A	A	A	S	S	A	S	A	S
A	A	S	S	A	S	A	S	A	S
A	A	S	S	S	A	A	S	A	A
S	S	A	A	S	A	S	A	A	S
A	S	S	S	A	S	A	A	S	A
S	S	S	A	A	S	A	S	A	S
A	S	A	S	S	A	S	A	A	S
S	A	S	A	S	S	A	S	S	A

Tabla 1.

El azar es sorprendente y los fenómenos aleatorios tienen una variabilidad mucho más grande de la que nos imaginamos.

de Jorge. ¿Podrías decir cuál de las dos es la tabla real y cuál es la inventada?

La segunda tabla contiene los resultados reales de un volado en tanto la primera contiene resultados falsos que tratan de imitar el azar. ¿Cómo se puede saber? Uno podría pensar que la primera debería ser la buena, porque en la segunda aparecen corridas o rachas muy largas de águilas consecutivas o de soles seguidos. La intuición nos dice que, si los resultados son al azar, no debería haber más de tres o cuatro águilas o soles seguidos en ningún momento. Sin embargo, cuando se lanzan 100 volados hay una probabilidad moderada de que aparezcan rachas hasta de seis águilas o soles seguidos. La probabilidad de una racha de cierta longitud se

A	S	S	S	A	S	A	S	S	S
S	S	S	A	S	S	S	S	S	S
A	A	A	S	A	S	S	S	A	S
S	A	S	A	A	S	S	S	A	S
S	S	A	A	S	S	A	S	S	A
S	A	A	S	S	S	S	A	S	S
A	A	A	A	A	A	S	A	A	A
A	S	A	S	A	S	A	A	S	S
A	A	S	S	S	S	A	S	A	S
A	S	S	A	A	A	A	A	A	A

Tabla 2.

puede calcular por medio de la estadística. Resulta que es bastante más alta de lo que uno podría pensar. Así, la tabla falsa es la que *no* contiene rachas largas. Puedes usar esta idea para impresionar a tus amigos: pídeles que, en tu ausencia, lancen una moneda 100 veces o que inventen los resultados, pero que no te digan qué hicieron. A partir de las tablas tú puedes adivinar fijándote en el largo de las cadenas consecutivas de águilas o soles.

La moraleja es que el azar es sorprendente y que los fenómenos aleatorios tienen una variabilidad mucho más grande de la que nos imaginamos. Por eso algunos acontecimientos que parecen poco probables —como un aguacero que inunda la ciudad— no lo son tanto.

Planear mejor

Otro ejemplo práctico y verdadero donde se muestra cómo se puede usar la estadística en aspectos cotidianos es el siguiente. En una pequeña tienda de abarrotes ubicada en el centro de la ciudad de Guanajuato se registró durante cuatro meses el número de clientes que entraban cada día de la semana. Esto lo hizo el dueño a petición de su hijo, quien estudiaba matemáticas en la universidad. Con esta información se pudo cuantificar la probabilidad de que lleguen más de 100 clientes un fin de semana, o menos de 50 un lunes, por ejemplo. Conocer los patrones

del número de clientes por día de la semana y a lo largo de los meses es muy importante para planear mejor las actividades de la tienda. Por ejemplo, si se sabe que la mayoría de los clientes compran refrescos y que los viernes y fines de semana llegan más personas, el dueño puede abastecer-

MÁS INFORMACIÓN

- <http://cuentame.inegi.org.mx/ayuda/estadistica.aspx?tema=A>
- <http://ponce.inter.edu/cremc/estadistica.htm>
- www.estadistica2013cimat.mx/que-es



se de refrescos suficientes para el fin de semana y así evitar que los clientes se vayan a la tienda de junto. También puede planear qué productos nuevos le conviene ofrecer, o saber si le conviene contratar un empleado adicional para los días de más afluencia.

Volvamos a las estadísticas del INEGI. La población económicamente activa de México en 2010 era de 50 693 306, de los cuales 31 314 268 eran hombres y 19 379 038, mujeres. Este dato es importante, pero no basta para muchos propósitos. Podríamos querer saber cuántos terminaron la primaria, cuántos cursaron estudios universitarios y de posgrado. Saberlo nos permitiría responder preguntas más interesantes y planear mejor aspectos de la educación y el empleo, pero estas estadísticas no hablan al respecto. Un montón de datos pueden dejar más preguntas que respuestas. Es importante estar preparado para leer estadísticas y aprender a analizarlas por medio de la estadística para sacar más provecho del azar. Lo más importante, como en ciencia en general y en matemáticas en particular, es pensar con cuidado. 🐼

Los autores agradecen a Mayra Nakamura por la preparación de las imágenes del tucán.

Ignacio Barradas es doctor en biología y matemáticas e investigador del Centro de Investigación en Matemáticas en Guanajuato. Ha colaborado en varias ocasiones con *¿Cómo ves?*

Elóisa Díaz-Francés es doctora en estadística e investigadora del Centro de Investigación en Matemáticas en Guanajuato.