¿Cómo sobrevivir a la incertidumbre?

Anabel Forte Deltell

NEXT— DOOR...
PUBLISHERS

© De la Autora:
Anabel Forte Deltell

© Next Door Publishers
Primera edición: septiembre 2022

ISBN: 978-84-125063-2-7
DEPÓSITO LEGAL: NA 1688-2022

Reservados todos los derechos. No está permitida la reproducción total o parcial de este libro, ni su tratamiento informático, ni la transmisión de ninguna forma o por cualquier medio, ya sea mecánico, electrónico, por fotocopia, por registro u otros medios, sin el permiso previo y por escrito de los titulares del *copyright*.

Next Door Publishers S.L.
c/ Emilio Arrieta, 5, entlo. dcha., 31002 Pamplona
Tel: 948 206 200
E-mail: info@nextdooreditores.com
www.nextdoorpublishers.com

Impreso por Estella Print
Impreso en España

Diseño de colección: Ex. Estudi
Autora del sciku: Laura Morrón Ruiz de Gordejuela
Editora: Laura Morrón Ruiz de Gordejuela
Corrección y composición: NEMO Edición y Comunicación, SL

El
Café
Cajal

A Adriana y Jorge, por ser la luz que ilumina mis días grises.
A Fran, por ser y por estar siempre.

Índice

Prólogo 13

Estadística —casi— por todas partes 19

1. Un lenguaje particular 25

2. No me representa 41

3. Todo lo medimos —y lo guardamos— 55

4. ¿Dónde bailan las variables discretas? 73

5. ¿Qué es normal? 91

6. Resumiendo, que es gerundio 105

7. La estadística de la prensa 129

8. De la parte al todo 167

9. Diagnóstico 189

10. ¿Quién será más alto? 213

11. Observar o experimentar 247

12. Paseos por un mar que crece 275

13. Todo cambia 303

Epílogo. Ciencia se escribe con R 315

Agra-
deci-
mien-
tos
325

Prólogo

Quejarse está de moda. O eso parece. Es bastante probable que mi amiga Anabel me regañase por hacer una afirmación de este tipo sin haber analizado un número suficiente y representativo de datos. O no. Ella no es de regañar, sino de enseñar con paciencia y dulzura.

Pero yo sigo pensando que quejarse está de moda. Y que pocas veces nos paramos, menos de las que debiéramos, a disfrutar y agradecer al azar las cosas buenas que nos encontramos por este camino maravilloso que es la vida.

Para mí, Anabel Forte es una de esas cosas buenas y bonitas que el azar y la vida han puesto en mi camino. Aunque, si lo examinamos y tenemos en cuenta que las dos somos matemáticas españolas y que nos gusta mucho Twitter, nuestro encuentro no ha sido, tal vez, fruto de la casualidad.

Desde que leí por primera vez una entrada en su blog, *BayesAna*, me quedé prendida de su estilo divulgando. Reconozco que echaba de menos o esperaba una voz como la suya contando matemáticas. Una voz seria, no aburrida, rigurosa y, sobre todo, dulce hablando de temas que pudieran resultar (para quienes no los conocen) fríos y deshumanizados como son la estadística. Anabel consiguió reconciliarme con esta rama de las matemáticas, tan maltratada en los medios y en los discursos políticos, y con la que yo nunca había tenido una relación demasiado amorosa. Por decirlo suave. Permítame por ello que comience este prólogo dándole a ella las gracias por ello.

Estoy segura de que este es el primer libro de muchos, desde los que su voz nos contará y enseñará mucha estadística a todos. Y falta que nos hace.

Y nada me hace más ilusión que el estreno editorial de esta gran divulgadora sea de la mano de una editorial tan comprometida con la difusión del conocimiento científico como esta de mis amigos de Next Door.

Dice un amigo mío que el 80 % de las personas que se quieren inventar un valor estadístico de algún dato dicen el 80 %.

Es una broma, pero me hace gracia y me pone nerviosa al mismo tiempo. La gracia está en que tratamos las probabilidades y las estadísticas con poca enjundia lo que nos lleva a equívocos en muchas ocasiones. Me pone nerviosa por exactamente el mismo motivo, una cosa es decir en un anuncio que el 80 % de los dentistas recomiendan tal o cual pasta de dientes y otra muy distinta cuando los datos estadísticos hacen referencia a cuestiones que pueden afectar gravemente a nuestras vidas.

Con la probabilidad tampoco lo llevamos mejor, no es fácil lidiar con ella. Simplemente somos malos, evolutivamente hablando, sopesando el significado de las probabilidades. Y, sin embargo, una vez más, son elementos esenciales para entender nuestro mundo. ¿Qué probabilidad hay de que llueva mañana? ¿Qué probabilidad tengo de tener un accidente en relación con mi edad, mi sexo o la cilindrada de mi coche? ¿Qué probabilidad hay de que no salga el Sol mañana? ¿Qué probabilidad tengo de sufrir una enfermedad si he dado positivo en un test? Todas estas preguntas, y muchas más, comparten una cuestión común que se puede resumir en: ¿qué sé del problema que me preocupa? Y de todas ellas se ocupa la probabilidad que es algo tan maravilloso que nos permite encapsular en números nuestro propio desconocimiento.

Por si fuese poco, usualmente, la probabilidad y la estadística quedan aparcadas a los últimos rincones de los libros de texto de secundaria o bachillerato. Esto hace que se lleguen a estos temas con prisa, si es que da tiempo tan siquiera. El problema está en la disposición y en la amplitud de los temarios, pero ese es un problema que no vamos a solucionar aquí.

Así que presentar este libro de mi amiga y compañera Anabel Forte me llena de felicidad. El libro es simplemente maravilloso. Anabel ha conseguido trazar un camino que nos lleva desde situaciones que todos hemos vivido a los conceptos matemáticos que hay tras la probabilidad y la estadística.

Se agradece tener un libro así entre las manos, un libro que aborda un tema complicado sin complejos, sin ser presuntuoso y sin ser paternalista. Anabel no duda en discutir cuestiones peliagudas con una naturalidad pasmosa. Sin pretender apabullar. Y eso se merece todo mi reconocimiento.

Como dice Anabel, la estadística y la probabilidad están en todas partes. Y en más que van a estar en este presente que no es otra cosa que ese futuro que se ha hecho mayor. Vivimos rodeados de datos, porcentajes, probabilidades, en las redes sociales, cuando vamos a contratar un seguro, cuando miramos nuestro móvil para decidir si tenemos que llevar paraguas. Pero es que ahora mismo nosotros somos fuentes andantes de datos, alimentamos sin control aplicaciones que siempre están hambrientas. Y esos datos son una parte muy íntima de lo que somos. Estamos condenados a tener un buen entendimiento de distribuciones de probabilidad, de inferencias estadísticas, de medias, medianas y modas. Tenemos que distinguir entre la interpolación y la extrapolación. Hemos de entender que la probabilidad es una medida de nuestro desconocimiento de un determinado fenómeno y cómo al actualizar nuestro conocimiento sobre dicho fenómeno puede cambiar las probabilidades que asignamos a tal o cual resultado. Podrías preguntarte, ¿y todo esto para qué? La respuesta es simple: para todo. Estamos viviendo en un mundo gobernado por el *big data*, por cantidades ingentes y casi insultantes de datos que tú y yo estamos empeñados en generar y, frecuentemente, en regalar. Y con esos datos, o, mejor dicho, con el análisis estadístico de esos datos es con los que tomamos decisiones y se toman decisiones que nos afectan.

Por todo ello, es una obligación entender y aprender cómo se obtienen y cómo se manejan datos y, sobre todo, cómo se analizan para obtener conclusiones y aplicaciones. Anabel ha conseguido construir un relato fabuloso donde todas estas cuestiones aparecen de forma natural. Este es un libro que se deja leer, que te exige atención, pero que te regala flores en forma de sorpresas, ideas y aventuras que te harán reflexionar y te maravillarán.

Desde mi propia experiencia puedo afirmar que da mucho vértigo enfrentarse a un papel en blanco o a una sala llena de gente mirándote y disponerte a contar matemáticas de forma que se puedan saborear. En esta ocasión, Anabel ha sabido hilvanar ideas y conceptos de una forma sutil y bonita. Este libro es una delicia para saborearla sin prisa, leyendo, pensando, releyendo, repensando.

Una de las cosas que más me gusta de este libro es que es para todo el mundo. Lo puedes leer desde cualquier escalón de la escalera del conocimiento matemático. Te va a servir, y mucho, si estás empezando con estos temas en el instituto porque podrás entender el porqué de cada escalón. Te va a servir si tienes que enseñar en secundaria y bachillerato porque aquí se describe el alma que habita en las fórmulas. Si estás en la parte alta de la escalera podrás disfrutar de la forma que tiene Anabel de describir un paisaje que conoces y podrás disfrutar de los detalles con facilidad. Y vosotros, amigas y amigos que estáis pensando en empezar a subir, no lo dudéis. La escalera es firme y os llevará hasta un lugar que os permitirá disfrutar de brisas nuevas y os descubrirá sitios desconocidos a los que os apetecerá volver.

Este libro se ha de leer tal y como se ha escrito, con amor, con paciencia y con muchas ganas. Es un libro de matemáticas, de probabilidades, de estadística, sí. Pero también es un libro en el que cada idea está habilidosamente imbricada con nuestro día a día. Es un lugar al que se viene a descubrir y a ser acompañados por una maravillosa anfitriona, una que conoce cada rincón de los lugares que nos va a descubrir con una pasión y una alegría de la que es difícil no contagiarse.

Y, como he dicho unas líneas antes, esta obra está acunada por una editorial que tiene un lugar especial en mi corazoncito. Next Door apuesta por una divulgación actual, una divulgación valiente que no evita las matemáticas, sino que las mima y les da el lugar que se merecen dentro del conocimiento científico. Por todo ello, para mí ha sido una gozada disfrutar de cada página de este libro. Porque me ha hecho sonreír, porque he recordado cosas que tenía en el cajón de la memoria, porque me ha hecho pensar y porque me ha hecho aprender. No es poca cosa esa para un libro que pretende popularizar contenidos matemáticos.

Escribir estas líneas ha supuesto todo un reto porque hay tanto que decir y tan poco espacio que no puedo expresar apropiadamente lo mucho que he disfrutado de la lectura de este libro. Espero que tú te encuentres en la misma situación al terminar. Eso sí, desde ya puedo afirmar que te va a encantar esta lectura. Sí, con probabilidad 1, con una certeza absoluta.

Pasen y lean.

Muchas gracias, Anabel.

Clara Isabel Grima Ruiz,
doctora en Matemáticas y profesora titular
de la Universidad de Sevilla

Estadística —casi— por todas partes

Recuerdo perfectamente el momento en el que supe que quería gritarle al mundo que la estadística estaba en todas partes.

Era una tarde, volvía a casa en el coche escuchando el programa de radio *Julia en la onda*. En ese momento, Julia Otero entrevistaba a una investigadora acerca de los *stents*, un tipo de malla elástica que se introduce en un vaso sanguíneo para ampliar el espacio de paso de la sangre y mejorar la circulación. No recuerdo los detalles, pero sí que describían un proceso de recogida de datos para comprobar que el uso de dicho mecanismo era efectivo, lo que, como veremos más adelante, se denomina «ensayo clínico».

Todo lo que se comentaba en la entrevista tenía una carga estadística enorme y, sin embargo, el término no llegó a aparecer nunca. Evidentemente, entiendo que así fuese, porque no era de lo que se estaba hablando y el tiempo en radio es limitado, pero, al mismo tiempo, resonaba en mi cabeza la idea de que lo que no se nombra no existe.

Y a la estadística le pasa un poco eso, está detrás de casi todo lo que nos es útil en nuestro día a día. Cuida de nuestra salud, de nuestras finanzas, de nuestras ciudades y, sin embargo, parece no existir, no estar ahí, porque nadie la nombra.

Pero espera…; céntrate, Anabel, que ya has usado bastantes términos raros y no has explicado nada. Empecemos desde el principio: ¿qué son «datos»?, ¿qué es «la estadística»? ¡Buenas preguntas!

En el caso particular de los datos, cuando hago esta pregunta en algún taller con alumnado de secundaria, la respuesta suele ser: «Eso que tenemos en el móvil»; y sí… pero no. Si nos fijamos en el origen de la palabra, *dato* proviene del latín *datum,* que significa 'lo que se da'. En ese sentido, un dato puede ser el número de teléfono de tu *crush*, la altura de una o varias personas o el nivel de polen en el aire de tu ciudad.

Si nos centramos algo más en el objetivo de este libro, podríamos decir que un *dato* es una observación, posiblemente distorsionada, de una realidad que queremos entender. Por ejemplo, en las mediciones de polen en el aire seguramente nos interesa saber cuál es ese nivel en cualquier lugar de la ciudad o incluso si es diferente entre los distintos barrios; para ello sería genial poder coger todo el aire y tener una medición precisa del polen, pero supongo que queda claro que no va a poder ser. Lo que sí podemos hacer es escoger una muestra de ese aire y medir el nivel con un aparato que, por muy preciso que sea, tendrá cierto error y, a partir de ella, extraer conclusiones sobre el aire en general, que es lo que solemos llamar «la población».

Sin embargo, extraer conclusiones sobre la población a partir de una muestra no es algo directo. Nos hace falta una herramienta que permita tener en cuenta que no lo hemos observado todo y que aquello que sí hemos observado tiene cierto error.

Profundicemos ahora en la segunda pregunta: ¿qué es la estadística? Si hacemos como con la palabra *dato* y nos vamos al origen del vocablo, *estadística* proviene del alemán *Statistik,* que significa 'la ciencia del Estado'. Este nombre se lo debemos —según parece— a Gottfried Achenwall (1719-1772), un historiador, filósofo y economista prusiano que, en su libro *Compendio de la constitución política de los países y pueblos europeos*, publicado en 1749, hizo referencia con esta denominación al análisis de los datos producidos por el Estado y sus habitantes. Es importante aclarar, sin embargo, que no sería justo reclamar que Achenwall sea el «padre» de la estadística —ni como término ni como ciencia—, ya que estamos hablando de una época en la que los datos empezaban a ser cada vez más importantes para el gobierno de los Estados y las ciudades, para entender la composición de estas, para saber cuánto tiempo vivían sus habitantes o de qué morían; y en este escenario había muchas mentes inquietas, además de Achenwall, tratando de encontrar herramientas que ayudasen en estos menesteres. Es importante destacar, por ejemplo, el papel de John Graunt (1620-1674), un mercero inglés que, cien años antes del trabajo de Achenwall, ya utilizaba técnicas que podríamos llamar «estadísticas» para estimar las tasas de mortalidad de la ciudad de Londres. También podríamos referirnos a Edmond Halley (1656-1742) —sí, el del cometa—, a quien debemos los primeros cálculos de la esperanza de vida con técnicas muy similares a las utilizadas hoy en día.

En definitiva, el término *estadística*, tal y como Gottfried lo utilizaba, comprendía diversas metodologías que servían para entender la población de un Estado y ayudar a su mejor gobierno con base en los datos de una muestra. Una serie de disciplinas que hoy englobamos dentro del término *demografía*.

Sin embargo, la estadística entendida como la ciencia que permite procesar, analizar e interpretar un conjunto de datos con el fin de poder extraer conclusiones o tomar decisiones tiene un origen mucho más antiguo. Por citar un ejemplo que me gusta en especial: Tucídides (460 a. C.-¿396? a. C.) cuenta en su *Historia de la guerra del Peloponeso* que en el 428 a. C., para determinar la altura de un muro durante el asedio a la antigua ciudad de Platea, se pedía a los soldados que contaran el número de ladrillos que tenía el muro desde la base al borde superior. Se seleccionaba entonces el valor más repetido entre los cálculos realizados —algo que en estadística llamamos «moda»— y se multiplicaba por la altura de un ladrillo[1]. A partir de este sencillo

1. Grattan-Guinness, I., *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, volumen 2, Routledge, 2004.

cálculo se aproximaba la altura deseada y se podían construir cuerdas que permitiesen escalar dicho muro.

La cuestión es que, desde estas primeras aproximaciones a la estadística hasta el uso que le damos en la actualidad, ha llovido mucho. La base teórica que la sustenta, las técnicas para la resolución de todo tipo de problemas y sus campos de aplicación han crecido de forma sorprendente hasta el punto de convertirla, junto con las matemáticas en general, en la disciplina que se encuentra tras la mayor parte del método científico. La ciencia detrás de las ciencias.

Cabe destacar, además, que los avances computacionales iniciados por la máquina analítica de Charles Babagge (1791-1871), los primeros programas de Ada Lovelace (1815-1852) y las ideas pioneras sobre inteligencia artificial de Alan Turing (1912-1954) han llevado la aplicación de la estadística a un nuevo nivel. Un nivel en el que la estadística y la ingeniería informática se entrelazan para dar lugar a lo que se ha llamado «ciencia de datos», una nueva disciplina cuya primera formalización se la debemos a William S. Cleveland (1943) en su artículo «Data Science: An Action Plan for Expanding the Technical Areas of the Field of Statistics»[2] y dentro de la cual se engloban esos términos tan recurridos como *machine learning* o *big data* que parecen ser la panacea computacional que permitirá solucionar todos los problemas pero que, sin una buena base estadística, pueden ser también el origen de todos los males.

Pero, bueno, lo cierto es que no he venido aquí a quejarme, que para eso ya tengo Twitter. Este libro es otra cosa. Es mi intento por visibilizar la gran cantidad de aspectos de nuestro día a día que se rigen por la estadística y cómo la necesitamos y la hemos necesitado siempre para entendernos; para entender nuestra salud, la prensa, las

2. Cleveland, W. S., Data science: «An action plan for expanding the technical areas of the field of statistics», *International Statistical Review*, 69(1), 2001, pp. 21-26.

decisiones políticas e incluso la aplicación del tiempo que llevamos instalada en el móvil.

Intentaré que sea un paseo; no sé si siempre lo conseguiré, pero prometo no abrumarte con muchas fórmulas ni todos los detalles técnicos, ya que mi propósito es que puedas captar la esencia y entender la potencia de la herramienta a través de la motivación y la historia. Así que no, no necesitas tener una formación previa concreta ni una calculadora a mano, solo dejarte llevar.

Aunque, antes de empezar y como quiero hablarte de nuestro día a día, creo que es mejor que nos pongamos en situación e intentemos vernos reflejados en personas como tú y como yo; bueno, quizás no sean exactamente como tú ni como yo. Tampoco serán como muchas otras personas, que las hay muy diversas y sería imposible representar a todo el mundo…

La cuestión es que a lo largo de este libro usaré la cotidianidad de una familia para entender la estadística que nos rodea. En concreto, nos acompañarán Lucía, Andrés, una pequeña terremoto de tres años llamada Alba y un aprendiz de caminante de casi doce meses, Mario.

Lucía y Andrés tienen suficiente con intentar conciliar y sobrevivir a la aventura de ser padres como para ser conscientes del impacto real que la estadística tiene en su día a día. ¿Que cuál es ese impacto? ¿Que a ti tampoco te parece para tanto? ¿Me dejas que te cuente?

«Estadística proviene del alemán *Statistik*, que significa ‘la ciencia del Estado’».

1

Un lenguaje particular

La voz de Àngels Barceló comenzó a sonar de pronto en el dormitorio. Era la señal de que el descanso nocturno se había acabado; bueno, si es que había existido… El pequeño Mario había pasado toda la noche gimoteando; algo le molestaba y Lucía se repetía una y otra vez: «Es muy probable que se despierte con fiebre».

Mientras en su cabeza rondaban esos pensamientos, Lucía remoloneaba y daba vueltas en la cama sin muchas ganas de levantarse; ya era martes, pero le seguía pareciendo lunes. Àngels seguía dando cuenta de las noticias en el radiodespertador que el hermano de Andrés les había regalado al poco de mudarse juntos, y que ella, a pesar de los años y el avance de las alarmas de los móviles, se resistía a desechar. Se autoengañaba pensando que era la mejor forma de no empezar el día frente a una pantalla; toda una quimera pues, tan pronto ponía un pie en el suelo, camino del cuarto de baño, lo primero a lo que su mano daba alcance era el móvil.

Se había imaginado viviendo en ese Estados Unidos de algún *blockbuster* de los noventa en el que su día hubiese comenzado con un «¡Buenos días, habitantes de Miami! ¡Levántense!, el sol brilla en un precioso cielo despejado». En cambio, sentada en el baño, Lucía consultaba en su móvil la aplicación del tiempo para planificar la ropa con la que llevar a Mario a la guarde y decidir si tenía que preparar las botas de agua para Alba.

«Probabilidad de lluvia del 50 %»

Otra vez la maldita probabilidad; pero ¿qué significaba eso de que hubiese un 50 % de probabilidad de lluvia? ¿Se iba a pasar lloviendo la mitad del día? Y la idea de que era probable que el peque se despertase con fiebre ¿qué narices quería decir? ¡Maldita incertidumbre!

Espera, vamos por partes: ¿probabilidad?, ¿incertidumbre? Estoy segura de que son conceptos que aparecen en tu mente con regularidad, como lo hacen ahora en la de Lucía y como lo han hecho en la del ser humano desde el inicio de los tiempos.

Orígenes

La incertidumbre, entendida como la falta de certeza que genera desazón o inquietud, es algo que nos ha acompañado siempre y que relacionamos sobre todo con los juegos de azar, a los que llevamos jugando desde la antigüedad. De hecho, tenemos constancia de que ya en el antiguo Egipto se jugaba a las tabas, pequeños huesos, normalmente del tobillo de una cabra, también llamados «astrágalos», que se utilizaban como precursores de los dados.

Por dar un dato curioso, la palabra *azar* tiene el mismo origen que *azahar* —la flor del naranjo— y significa 'flor'. Algunos textos afirman que esto se debe a que, para identificar la cara ganadora de las tabas, se dibujaba en ella una flor.

La cuestión es que en el contexto de este tipo de juegos y sus consiguientes apuestas surge la necesidad de cuantificar la incertidumbre, comenzando así los primeros intentos para comprenderla y dominarla. Sin embargo, no es hasta el siglo XVI cuando aparecen los primeros estudios formales sobre el azar y aún habría que esperar cien años más para que Blaise Pascal (1623-1662) y Pierre de Fermat (1601-1665) sentaran las bases de la teoría de la probabilidad tal y como la conocemos hoy.

Soy consciente de que todo esto que te acabo de contar, más allá de ayudarte a ganar algún quesito en el Trivial, no te ha permitido entender qué es eso de la probabilidad, así que vamos a entrar en materia.

Como comentaba, la teoría de la probabilidad surge ante la necesidad de medir la incertidumbre. Sin embargo, para poder adentrarnos en ella primero debemos entender a qué nos referimos cuando decimos que algo es incierto o, más concretamente, cuando hablamos de «aleatoriedad».

En estadística decimos que algo es aleatorio cuando no podemos conocer su resultado antes de observarlo. Esta idea es fácilmente aplicable al lanzamiento de una moneda en el que no puedes prever si saldrá cara o cruz. Aunque también puede aplicarse a muchas otras situaciones cotidianas, como la fiebre de Mario. Lucía puede intuir que será superior a 36 grados con un beso en la frente, pero hasta que no le ponga el termómetro la temperatura real será desconocida, incierta y, por tanto, nos encontramos ante un suceso aleatorio. Algo similar pasa con la lluvia. Saber si lloverá de camino al cole será una incógnita hasta el momento de ir al cole.

«La teoría de la probabilidad surge ante la necesidad de medir la incertidumbre».

Hay que mencionar que existen ciertas corrientes de pensamiento, como la defendida por Bruno de Finetti (1906-1985), en las que se considera que la aleatoriedad que percibimos es solo un reflejo de nuestra falibilidad como seres humanos, de nuestra capacidad reducida para conocer el mundo. En ese sentido, podría ser que la única aleatoriedad real residiera en niveles subatómicos. Es a esos niveles donde nos encontramos con el «principio de incertidumbre» enunciado por Werner Heisenberg (1901-1976), en el que se demuestra que no es posible determinar dos constantes físicas complementarias, como la posición y la cantidad de movimiento de una partícula, con una precisión concreta.

Por supuesto, a Lucía esto le trae bastante al pairo; ahora mismo, sentada en el baño, quiere saber con total exactitud si va a llover, pero el caso es que la aplicación del móvil sigue mostrando el mismo mensaje: «Probabilidad de lluvia del 50 %», así que no le queda otro remedio que intentar entender qué significa. Y no es fácil.

Interpretar la probabilidad

Jenny Gage y David Spiegelhalter, en su libro *Teaching Probability*[3], alertan de la dificultad de la población general para entender la forma en que transmitimos la probabilidad y proponen que no debería expresarse como porcentaje entre 0 y 100, ni como un valor entre 0 y 1, como es habitual en el mundo de las matemáticas y la estadística, sino de una forma que resulte más intuitiva mediante el uso de metáforas o símiles. En concreto, Gage y Spiegelhalter proponen expresar que algo tiene una probabilidad de 0,5 —o del 50 %— indicando que, de cada 100 veces que se repita la misma situación, en 50 de ellas ese algo sucederá. De hecho, en el caso de la lluvia que amenaza el día de Lucía y Andrés, la forma de calcular la probabilidad tiene exactamente ese sentido.

En concreto, lo que se hace para determinar la probabilidad de que caiga un chaparrón, *grosso modo*, es utilizar una especie de simulador, un modelo —de estos te hablo en el capítulo 12— que, a partir de ciertas condiciones, como pueden ser la temperatura, la presión o la humedad, te dice si va a llover o no. Este tipo de simuladores tienen la particularidad de que, al repetir el proceso con los mismos datos, el resultado no siempre es igual —por muchas cosas que ya

3. Gage, J. y Spiegelhalter, D., *Teaching Probability*, Cambridge University Press, 2016.

iremos viendo—. Así pues, lo que se hace es simular muchas veces con condiciones similares y ver cuántas de las simulaciones terminan en lluvia. De este modo, una probabilidad de lluvia de un 50 % nos dice que, de 100 simulaciones, en 50 llovió y en otras 50 no. Esto es como decir que es mejor echar una muda extra y las botas de agua en una bolsa porque no está nada claro si lloverá o no.

Si la probabilidad hubiese sido del 10 %, lo que tendríamos que entender es que, de cada 100 simulaciones, solo 10 indican que lloverá. Incluso, si quieres simplificar la interpretación, podrías pensar que, de 100 días iguales al de hoy, solo llovería en 10 de ellos; por lo tanto, si nos dejamos las botas de agua en casa, tampoco arriesgamos tanto.

Y lo cierto es que el uso de la expresión «tantos de cada cien» puede resultar bastante intuitivo; sin embargo, el lenguaje que se usa en los medios de comunicación para cuantificar un riesgo no siempre lo es tanto.

Es habitual, por ejemplo, que al querer expresar una probabilidad que está por debajo de «1 vez de cada 100» tendamos a hablar de una cierta cantidad de cada 1000 o de cada 10 000 así, comparar los riesgos entre sí se vuelve algo más confuso, pues ¿qué riesgo es mayor: 1 de cada 1000, o 1 de cada 10 000? Os sorprendería saber la cantidad de veces que nos equivocamos al interpretarlo, ya que nuestra mente no termina de procesar de manera correcta que un riesgo menor se expresa haciendo uso de un valor numérico mayor —10 000 frente a 1000, por ejemplo—. Para poder abstraernos y entenderlo un poco mejor os invito a imaginar un bombo de la lotería donde introducimos un número de bolas negras y una única bola roja. Piénsalo: ¿cuándo será menor el riesgo de que salga la roja, cuándo haya 99 bolas negras —1 roja de 100— o cuando haya 9999 —1 roja de 10 000—? Ahora más fácil, ¿verdad?

Si además nos planteamos cómo se calcula una probabilidad, la cosa se complica. Por una parte, podemos pensar en la conocida como «regla de Laplace» en honor al matemático francés Pierre-Simon Laplace (1749-1827). Mediante esta, la probabilidad de que suceda algo se calcularía como la división del número de situaciones que dan lugar a ese algo entre el número total de posibles situaciones. Esto es sencillo cuando hablamos de, por ejemplo, el lanzamiento de un dado: la probabilidad de que salga un número par es de 3/6 —3 soluciones favorables: el número 2, el 4 y el 6 de 6 posibles—; o lo que es lo mismo: 0,5 o 50 de cada 100.

Sin embargo, si volvemos a la probabilidad de que Mario tenga fiebre, pensar de esta forma nos llevaría a la errónea conclusión de: tengo dos posibles situaciones —fiebre vs. no fiebre—, de las cuales una da lugar a la situación que buscamos —fiebre—. Si usamos la regla de Laplace, el resultado sería de 1/2, 0,5, pero ya intuimos que se trata de una probabilidad mayor. Y es que la regla de Laplace solo funciona cuando todas las posibilidades involucradas tienen exactamente la misma probabilidad de suceder. Y esto se vuelve un poco la pescadilla que se muerde la cola porque para poder calcular de esta forma la probabilidad de que Mario tenga fiebre deberíamos asegurarnos de que la probabilidad de que tenga fiebre es la misma que la de que no la tenga. Es decir, para calcular la probabilidad de que tenga fiebre debemos conocer la probabilidad de que tenga fiebre; un poco raro, ¿no?

Otra opción implica pensar en la probabilidad de que algo suceda como la frecuencia de veces que realmente sucederá si repetimos la misma situación infinitas veces, y esto puede tener sentido si pensamos en lanzar una moneda miles de veces bajo las mismas condiciones —lo que sea que signifique «mismas condiciones»— y calcular la frecuencia con la que sale cara. Sin embargo, para entender así la probabilidad de que Mario tenga fiebre tendríamos que pensar en repetir la noche de hoy infinitas veces, como en la película *Atrapado en el tiempo*, y averiguar en qué proporción de ellas Mario tiene fiebre, pero, como entenderéis, a Lucía se le hace cuesta arriba.

Por último, existe la posibilidad de pensar en la probabilidad como una medida personal de lo verosímil o creíble que es que algo suceda, y que puede estar basada en un conocimiento previo sobre el tema, y aquí las observaciones de Lucía sobre lo mal que Mario ha pasado la noche tienen mucho que decir.

Esta forma de calcular la probabilidad engloba, en cierta forma, a las dos anteriores, ya que un cálculo de lo verosímil que resulta algo, basado en la frecuencia o en la regla de Laplace, no deja de ser una medida de su incertidumbre basada en el conocimiento que se tiene sobre el problema.

En cualquier caso, sea cual sea la interpretación que hagamos de la probabilidad, la forma que tengamos de comunicarla o calcularla, es importante recordar que se trata de una entidad matemática y que, como tal, existen reglas, denominadas «axiomas», sobre cómo debe ser y qué propiedades debe cumplir.

Axiomas: las cosas claras

No os asustéis por el palabro; un axioma en matemáticas es una norma tan básica que no hace falta demostrarla. Un ejemplo de axioma es que, si dibujo dos puntos en un folio —un plano, matemáticamente hablando—, solo existe una recta que pase por los dos a la vez; venga, a ver quién es tan valiente como para decir que no es cierto.

En el caso de la probabilidad, los axiomas que le dan base y desde los que partimos son relativamente recientes, en concreto se los debemos a Andréi Nikoláyevich Kolmogórov (1903-1987) y vieron la luz en 1933.

El primero de estos axiomas o reglas establece que la probabilidad debe ser un valor entre 0 y 1, o entre el 0 % y el 100 %, donde el 0 % es la probabilidad de algo que no puede pasar, que es imposible —bueno… imposible con matices, pero eso te lo cuento en el capítulo 5—, y 100 % es la probabilidad de algo que pasará seguro, sí o sí, sobre lo que no hay incertidumbre.

La siguiente regla nos dice que, cuando dos eventos no pueden suceder a la vez, la probabilidad de que suceda alguno de ellos será la suma de la probabilidad de cada uno; me explico: supongamos que Alba, la hija mayor de Lucía y Andrés, elige ponerse sus zapatillas azules con una probabilidad del 40%, —40 de cada 100 veces— y las verdes, del 20% —20 de cada 100 veces—. Pues bien, la probabilidad de que esa mañana acabe saliendo de casa con zapatillas azules «o» verdes será del 60%. La clave de esta regla está en que no puede ponerse las azules y las verdes a la vez —y no me vengáis con que se puede poner una de cada, que esto es solo un ejemplo y os veo venir—. Siguiendo con el mismo ejemplo, los axiomas de Kolmogórov nos dicen que la probabilidad de que salga a la calle sin zapatillas es del 0% —suceso imposible, que todos hemos sido pequeños a cargo de personas adultas y nos sabemos las reglas— y, por descarte, que se ponga algunas que no sean azules ni verdes, de un 40%, porque, al final, la suma de todo o, lo que es lo mismo, la probabilidad de que salga con zapatillas a la calle —suceso seguro— debe ser del 100%.

A partir de estas reglas se puede operar matemáticamente con la probabilidad e ir complicando el escenario para calcular, por ejemplo, lo que se llaman «probabilidades condicionadas». Esto podría ser la probabilidad de que Mario tenga fiebre bajo la condición de haber pasado una mala noche, que seguro será mayor que la probabilidad de que tenga fiebre después de una buena noche, porque las cosas que nos pasan no son siempre independientes.

Probabilidad condicionada

En ese sentido decimos que dos eventos son independientes cuando la probabilidad de que se den ambos se puede obtener como el producto de la probabilidad de que pase cada uno individualmente. El ejemplo más sencillo lo tenemos al pensar en la probabilidad de que salgan dos caras seguidas en el lanzamiento de una moneda. La primera vez la probabilidad es 0,5, la segunda también —sin importar lo que haya pasado en la primera—, pero la probabilidad de que ambas sean cara será el producto de: $0{,}5 \times 0{,}5 = 0{,}25$.

La cuestión es que, como comentábamos en el caso de la fiebre de Mario, si dos sucesos no son independientes, al saber que ha pasado uno, tendremos información que nos cambiará la probabilidad de que haya pasado el otro.

Llegados a este punto es muy importante estar siempre atentos a cuál es el evento sobre el que queremos calcular la probabilidad. ¿Que por qué digo esto? Pues porque no es lo mismo calcular la probabilidad de que Mario tenga fiebre sabiendo que ha pasado mala noche que la probabilidad de que Mario haya pasado mala noche sabiendo que ha tenido fiebre. Si nos fijamos bien, en el primer caso, el suceso que nos interesa es si Mario tenía fiebre o no sabiendo —porque ya ha pasado— que ha estado molesto toda la noche, mientras que, en el segundo caso, nos preocuparía el hecho de que duerma mal al saber que tiene fiebre, una preocupación más típica de antes de acostarnos que de cuando acabamos de levantarnos.

Esta dualidad puede parecer obvia en este caso, pero lo cierto es que es fácil confundirse y, de hecho, ha sucedido muchas veces a lo largo de la historia, sobre todo en el ámbito de la justicia. Por este motivo, esta interpretación errónea ha acabado conociéndose como la «falacia del fiscal» e implica que se confunda la probabilidad de que se encuentren ciertas pruebas asumiendo que la persona acusada ha cometido el crimen con la probabilidad de que esa persona sea culpable en presencia de las evidencias.

Uno de los casos más famosos en los que esta falacia llevó a la cárcel —alerta, *spoiler*— de manera injusta a una persona fue en el juicio contra Sally Clark[4]. Sally estaba acusada de haber asesinado a

4. Shaikh, T. «Sally Clark, mother wrongly convicted of killing her sons, found dead at home», *The Guardian*, 2007. Disponible en:

sus dos hijos pequeños, que habían fallecido por muerte súbita, un suceso muy triste pero que puede producirse, de forma natural, en los primeros meses de vida de algunos bebés.

Durante el juicio se incurrió en la falacia del fiscal al considerar lo probable que era que los niños hubiesen muerto si la madre era la culpable, pero no la probabilidad de que la madre fuese culpable, un valor que se veía considerablemente reducido si se tenían en cuenta todas las posibles causas de muerte y sobre todo que la muerte súbita podía tener que ver con la genética de los bebés y, por tanto, ser más común entre hermanos.

Otro ejemplo de esta falacia y que encontramos en nuestra vida cotidiana es la probabilidad de tener cierta enfermedad habiendo dado positivo en la correspondiente prueba. Pongámonos en situación con un ejemplo real.

Cuando nace un bebé, una de las primeras pruebas que se le hacen es la del talón. Dicha prueba es un cribado, es decir, se realiza a todo el mundo —en este caso, a todos los bebés—, sin tener sintomatología previa. En ella se extrae una pequeña cantidad de sangre del talón —de ahí su nombre—, que se analiza con el fin de descartar algunas enfermedades raras y muy graves que pueden beneficiarse de una atención temprana.

Estas pruebas suelen tener una alta sensibilidad, es decir, una alta probabilidad de dar positivo si la persona sufre la enfermedad. Por ejemplo, en el caso del hipotiroidismo congénito (HC), la sensibilidad del cribado y posterior análisis es de un 97,5 %. Pues bien, imagina que le hacen esta prueba a tu bebé y da positivo.

Si caemos en la falacia del fiscal, nos asustaremos, pues parecería que la probabilidad de que nuestro bebé esté enfermo es del 97,5 %, pero no es cierto. Recuerda que ese valor es la probabilidad de dar

---

https://www.theguardian.com/society/2007/mar/17/childrensservices.uknews. Consultado el 21 de febrero de 2022.

positivo si está enfermo, no de estar enfermo. Para poder saber cuál es la probabilidad real de estar enfermo necesitamos algunos datos más y una fórmula conocida como «teorema de Bayes».

En concreto, necesitamos conocer la incidencia de la enfermedad en la población o, lo que es lo mismo, la probabilidad de que una persona al azar la sufra. En el caso del HC, estamos ante una incidencia de, aproximadamente, 1 persona de cada 2500; esto es, una probabilidad de 0,0004, un 0,04%.

También necesitaremos conocer la especificidad, la probabilidad de dar negativo cuando no se está enfermo y que en el caso de la prueba del talón para el HC es de un 99%[5].

Una vez establecidos estos dos valores, llega el momento de introducir la fórmula de la que os hablaba: el teorema de Bayes.

En busca de las causas: el teorema de Bayes

Este teorema se lo debemos al reverendo presbiteriano Thomas Bayes (1702-1761), quien, buscando entender cuáles eran las causas de las cosas que nos pasaban —véase, buscando entender el papel que jugaba Dios en todo este lío—, desarrolló un teorema que buscaba justamente pasar de la probabilidad de que algo ocurriese si se sabía la causa a entender cuál era la probabilidad de que algo fuese la causa de lo que pasaba.

El trabajo de Bayes fue publicado de forma póstuma en 1763 por su gran amigo Richard Price, que lo encontró entre muchos otros papeles mientras hacía limpieza en la casa de Thomas. Sin embargo, como pasa muchas veces en ciencia, Bayes no fue el único en llegar

5. Fernández Grande, «E. Cribado neonatal ampliado: "Prueba del talón"». Documento pdf, Hospital General Universitario de Ciudad Real, Ciudad Real, 2015. Obtenido de http://www.hgucr.es/wp-content/uploads/2015/05/CRIBADO-sinc.pdf. Consultado el 21 de febrero de 2022.

a esta fórmula; de hecho, la versión que utilizamos hoy en día se la debemos a Pierre-Simon Laplace, que, al tratar de demostrar que la ley de la gravitación de Newton era la causa de, entre otras cosas, el movimiento de los planetas llegó a las mismas conclusiones que el reverendo.

Pero vamos al lío: el teorema de Bayes en su forma más conocida nos habla de dos posibles sucesos a los que llamaremos A —tener HC— y B —dar positivo en la prueba del talón para HC—. Tenemos entonces la probabilidad de tener HC, que podemos representarla como P(A), y la probabilidad de dar positivo en la prueba del talón para HC —lo tengas o no—, que la escribiremos como P(B). Si lo que queremos conocer es la probabilidad de tener HC dado que la prueba ha dado positivo, lo que buscamos es la probabilidad condicionada que solemos representar por P(A|B), donde la línea vertical separa el suceso de interés de la condición. De la misma forma, P(B|A) es la forma de escribir matemáticamente la probabilidad de que pase B sabiendo que ha pasado A o, siguiendo el ejemplo, la probabilidad de dar positivo dado que se tiene la enfermedad.

Si usamos esta forma de escribirlo, el teorema de Bayes nos queda como se puede ver en el neón de la figura 1.

Figura 1. Neón mostrando el teorema de Bayes en la Universidad de Cambridge.

Volviendo a nuestro ejemplo, y como ya habíamos dicho, sabemos que la probabilidad de tener la enfermedad en general, es decir, la incidencia, era: $P(A) = 0{,}0004$. Por otra parte, sabemos que la probabilidad de dar positivo si se está enfermo es lo que antes hemos llamado «sensibilidad» y que, en este caso, es de un 97,5 %; esto es: $P(\text{positivo}|\text{enfermedad}) = P(B|A) = 0{,}975$.

Por último, nos faltaría la probabilidad que aparece en el denominador de la fórmula, P(B), que, como hemos comentado, en este caso es la de dar positivo así en general, sin saber si se tiene la enfermedad o no. Esta probabilidad, si bien no la sabemos directamente, se puede calcular mediante otro teorema, el de la probabilidad total.

El teorema de la probabilidad total nos viene a decir que, si somos capaces de dividir la realidad en diferentes escenarios y conocemos lo probable que es nuestro evento de interés en cada uno de ellos, podemos acabar sabiendo cómo es de probable ese algo en general. Esto, en términos matemáticos sería algo así:

$$P(\text{suceso}) = P(\text{suceso}|\text{escenario 1}) \times P(\text{escenario 1}) + \\ P(\text{suceso}|\text{escenario 2}) \times P(\text{escenario 2}) + \ldots$$

Volviendo a la prueba del talón, yo puedo dividir la realidad entre el bebé tiene la enfermedad o el bebé no tiene la enfermedad, ¿cierto?; no hay más opciones. Pues bien, yo sé cómo de probable es dar positivo en el primer caso: $P(\text{positivo}|\text{enfermedad}) = 0{,}975$. También sé cómo de probable es este primer escenario: $P(\text{enfermedad}) = 0{,}0004$. Perfecto. Ahora vamos con la segunda parte: ¿cuál es la probabilidad de dar positivo si no se tiene la enfermedad? Pues así directamente no lo sé, pero tenemos la especificidad que habíamos dicho, que era la probabilidad de dar negativo si no se está enfermo. Y como dar positivo y dar negativo son sucesos complementarios —o se da uno o se da el otro—, sus probabilidades deben sumar 1. Es decir, la probabilidad de dar positivo si no se está enfermo será uno menos la especificidad: $1 - 0{,}99 = 0{,}01$. Además, también conocemos cómo de probable es este escenario, ya que, si la probabilidad de tener la enfermedad es 0,0004, la de no tener la enfermedad será de 0,9996.

Y así, si lo juntamos todo y utilizamos la fórmula del teorema de la probabilidad total que acabamos de ver, tenemos que:

$$P(\text{positivo}) = P(\text{positivo}|\text{enfermedad}) \times P(\text{enfermedad}) + P(\text{positivo}|\text{no enfermedad}) \times P(\text{no enfermedad})$$

Pues bien, volvamos a los números:

$P(\text{enfermedad}) = 0{,}0004$
$P(\text{no enfermedad}) = 0{,}9996$
$P(\text{positivos}|\text{enfermedad}) = 0{,}975$; la sensibilidad
$P(\text{positivo}|\text{no enfermedad}) = 1 - \text{especificidad} = 1 - 0{,}99 = 0{,}01$
$P(\text{positivo}) = 0{,}975 \times 0{,}0004 + 0{,}01 \times 0{,}9996 = 0{,}0104$

Y ahora solo nos queda usar el teorema de Bayes sustituyendo cada elemento por su valor y obtenemos que la probabilidad de que nuestro bebé esté enfermo de HC tras un positivo en la prueba del talón es de aproximadamente 0,04; es decir, alrededor de un 4% de los bebés cuya prueba ha sido positiva tendrán HC. Un valor muy bajo que no tiene nada que ver con la seguridad que en un inicio nos alarmó.

Es posible que entre tanto número y fórmula no sea fácil entender esta aparente paradoja, pero vamos a plantearlo de forma más visual.

Piensa que le hacen la prueba del talón a 10 000 bebés. De esos 10 000, y dado que la enfermedad la tiene solo un 0,04%, tenemos que habrá solo 4 bebés con HC, mientras que 9996 bebés no tendrán la enfermedad. De esos 9996 bebés, la especificidad nos dice que un 99% tendrá un valor negativo, pero un 1% tendrá un positivo. Esto es, alrededor de 99 bebés que no tienen la enfermedad darán positivo.

Y de los 4 bebés que sí tienen la enfermedad, la prueba acierta en un 97,5 % de las veces, pero se puede equivocar en un 2,5 %. Vayamos al caso extremo, al caso en que no se equivoca nunca. Como vemos en la figura 2, con esto tenemos que el número de bebés que habrán dado positivo son: 4 + 99 = 103. Es decir, tengo 103 positivos de los cuales solo 4 tienen la enfermedad. Lo que quiere decir que sobre un 4 % de los bebés que den positivo tendrán la enfermedad o, dicho de otra forma, que la probabilidad de tener la enfermedad habiendo dado positivo es 0,04, tal y como habíamos obtenido con el teorema de Bayes.

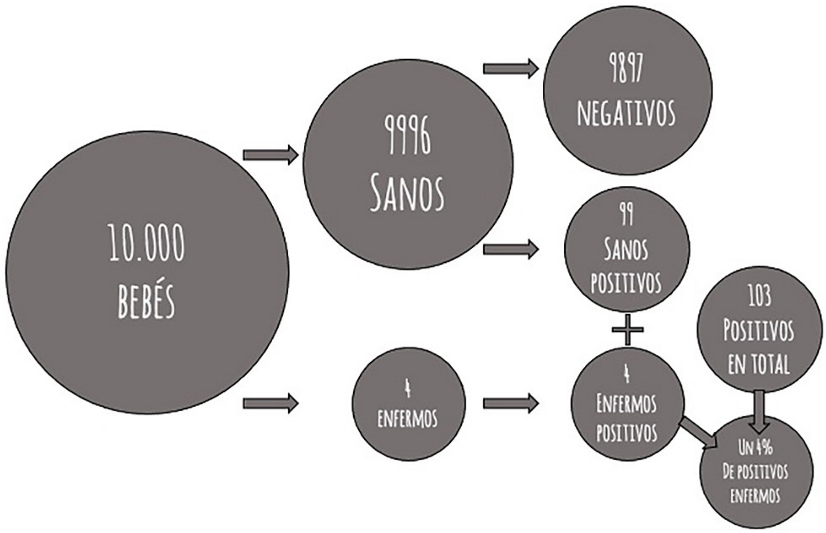

Figura 2. Representación gráfica del número de positivos/negativos y enfermos/sanos en diez mil bebés.

Por supuesto, es una probabilidad 100 veces mayor que la de tener la enfermedad si no has dado positivo, pero, aun así, mucho menor que la sensibilidad de la prueba.

Enredada en sus pensamientos sobre la probabilidad y el día que se avecinaba, Lucía empezó a escuchar cómo Andrés se había despertado también y le susurraba al bebé, que había pasado del sutil gimoteo a ese llanto dolorido que augura un lastimero «hoy nada de guarde».

Se apresuró a salir del baño para encontrarse con un termómetro que marcaba 38 grados; nada que hacer, se había despejado la incertidumbre, tocaba llamar al trabajo, organizarse con Andrés y llevar al peque al pediatra.