La incertidumbre atraviesa nuestra vida y para cuantificarla usamos la probabilidad y la estadística. Esta disciplina aprende de los datos y da estimaciones con intervalos que expresan su precisión lo que nos ayuda, no a eliminar la incertidumbre, pero si a medirla para tomar mejores decisiones.

[ILUSTRACIÓN: AJIJCHAN/ ISTOCK]

Es irremediable, nos inquieta no saber. Es una sensación intrínseca al ser humano que suele poner el foco en el futuro –qué pasará mañana–. Pero la incertidumbre no vive solo en el calendario. También se instala en los porqués, en la comprensión de los procesos que desembocan en lo que sucede: ¿por qué mido 1,65 metros? ¿Qué nos lleva –o no– a desarrollar un cáncer? ¿Qué genes influyen en cierta enfermedad? No es solo preguntarse si, al lanzar una moneda, saldrá cara, sino qué mecanismos físicos y contextuales empujan ese resultado y en qué medida lo hacen.

La realidad es compleja, ruidosa y a veces caprichosa; y nuestro conocimiento es inevitablemente limitado

No hay bola de cristal: la realidad es compleja, ruidosa y a veces caprichosa; y nuestro conocimiento es inevitablemente limitado. Lo que sí tenemos es un lenguaje para cuantificar esa falta de certeza y convertirla en algo con lo que podamos tomar decisiones: la probabilidad. Y contamos con una disciplina que ancla ese lenguaje a lo que observamos y medimos: la estadística.

La probabilidad de lo que nos preocupa

Quizás aterrizar la idea en un ejemplo nos ayude a entenderlo mejor. Pensemos en el cáncer de pulmón. La preocupación personal sobre la presencia de esta enfermedad se formula en términos binarios –"¿me tocará o no?"–, pero la respuesta a esta duda razonable la encontramos en la probabilidad. No es lo mismo que nos digan que la probabilidad de que desarrolles esta dolencia es de un 3 %, de un 20 % o de un 80 %, ¿cierto? Ninguna cifra garantiza el desenlace final, pero cambiará las decisiones que tomemos respecto a la frecuencia con la que nos haremos revisiones, nuestros hábitos y nuestras prioridades.

La cuestión está entonces en cómo determinar esa probabilidad. Podemos centrarnos en la búsqueda de estimaciones construidas con datos de personas con condiciones parecidas de edad, exposición ambiental, hábitos, antecedentes, etc., sin perder de vista que se tratará de eso, de una estimación y, como tal, tendrá asociada, de nuevo, cierta incertidumbre. Así pues, un estudio serio no afirmará "tu riesgo es de un 3 %" como si fuera una propiedad revelada, sino, "dadas tus características y los datos disponibles, el riesgo está muy probablemente entre un 1 % y un 5 %". Además, ese "muy probablemente" también puede cuantificarse, tratando de poner rango y coherencia a lo que sabemos y a lo que no.

¿Y de qué depende la calidad de esa estimación? Pues depende básicamente de tres pilares. Primero los datos: no se trata solo de cuántos, sino también de cuán bien representan a la población que nos importa. Segundo, el diseño: cómo se recogen esos datos, con qué controles, con qué cuidado para evitar sesgos. Tercero, el modelo: la simplificación matemática y las técnicas estadísticas que usamos para realizar la estimación y de cuya aplicabilidad dependerá obtener mejores o peores resultados con los mismos datos. Unos datos suficientes, obtenidos con un buen diseño y analizados con la técnica correcta nos llevarán

siempre un paso más hacia el éxito.

Ver que dos cosas se mueven juntas no basta para concluir que una causa la otra

A partir de aquí, asoma otra pregunta: qué variables están realmente implicadas en el riesgo. Volviendo al cáncer de pulmón, podemos valorar factores como tabaco, contaminación, ocupación, actividad física… Pero, ojo, porque ver que dos cosas se mueven juntas no basta para concluir que una causa la otra. Es posible que un gran estudio observacional encuentre que quienes hacen más deporte tienen menos cáncer de pulmón, pero quizá el vínculo real sea que quienes hacen más deporte no fuman. Es lo que, en estadística, llamamos confusión. Para despejarla, necesitamos trabajar en ese segundo pilar del que hablábamos, diseños más específicos y exigentes nos permitirán elegir mejor que variables incluir en el modelo o qué técnicas estadísticas serán más efectivas para dar una mejor respuesta con una mejor medida de la incertidumbre.

¿Cuánto me afecta?

Superada esa fase, la siguiente cuestión es la magnitud. No basta con saber si un factor influye, también necesitamos saber cuánto cambia el riesgo. Aquí conviene distinguir entre riesgo absoluto y relativo. Decir que "fumar duplica el riesgo" es hablar en términos relativos: si una persona no fumadora tiene 10 % de riesgo, duplicarlo lleva al 20 %. Suena igual de contundente cuando el riesgo pasa de 0,5 % a 1 %, pero el impacto práctico es distinto.

En este sentido, al hacer referencia al efecto de tal o cual hábito, puede que, en los medios de comunicación, hablen de riesgos relativos, de probabilidades o de ratio de probabilidades, pero lo verdaderamente importante es que esa jerga acabe convertida en números que quien lo lea pueda situar en su vida.

Dar un valor es fácil. Interpretarlo y entender las reglas que lo rigen, ya no tanto

Y una última obviedad que nunca sobra: al ver escrito uno de estos valores, puede que te presenten también un porcentaje. Es importante no confundirlo con una probabilidad y, en ese sentido, la probabilidad nunca supera el 100 %. Dar un valor es fácil. Interpretarlo y entender las reglas que lo rigen, ya no tanto.

Breve guía para entender la probabilidad

Volvamos al momento en que te informan de tu probabilidad de sufrir cáncer de pulmón. Lo habitual es verla expresada en porcentajes (entre 0 % y 100 %). También puede expresarse en tanto por uno —3 % es 0,03; 20 % es 0,2— o como "x de cada N" donde, por ejemplo, uno de cada 1.000 equivaldría a 0,1 % (0,001) o tres de cada diez, a un 30 %.

El problema surge cuando, en un mismo texto, se mezclan formatos, ya que tendemos a fijarnos más en el total (N) que en la parte (x) y podemos creer que "uno de cada 1000" supone un riesgo mayor que "uno de cada 100", cuando en realidad estamos hablando de un 0,1 % frente a un 1 %.

Entendido el número, toca interpretarlo y, para ello, contamos con dos miradas diferentes. La mirada frecuentista entiende la probabilidad como la frecuencia con que veríamos un evento al repetir el proceso muchas veces. Funciona muy bien en contextos controlables y repetibles. Pero, al hablar de la probabilidad de lluvia o de que una persona enferme –donde no podemos "repetir" el mismo día o la misma vida–, resulta más natural la mirada bayesiana, que trata la probabilidad como un grado de credibilidad coherente y actualizable con nueva evidencia mediante el teorema de Bayes.

Estas dos perspectivas también ayudan a entender la incertidumbre de las estimaciones. Los rangos de valores plausibles de los que hablábamos antes reciben el nombre de intervalos de confianza bajo el paradigma frecuentista o de intervalos de credibilidad bajo el bayesiano. En ambos casos, nos ayudan a medir la fiabilidad de la estimación. No es lo mismo un riesgo estimado entre 1% y 5 % que entre 1 % y 10 %. Quizá ambas cifras nos dejen igual de tranquilos ante la enfermedad, pero el primer intervalo es más preciso y, por tanto, más informativo sobre lo que sabemos e indica, posiblemente, mejores datos o un mejor uso de estos.

La incertidumbre está presente en cualquier proceso cotidiano y no es posible eliminarla

En definitiva, la incertidumbre está presente en cualquier proceso cotidiano y no es posible eliminarla, pero sí medirla. Al hacerlo la volvemos visible y honesta. Probabilidad y estadística no prometen certezas, pero sí mejores preguntas y decisiones más conscientes. No se trata de controlar un bosque lleno de rincones oscuros sino de caminar por él con mejor criterio.