

Los siete pilares de la sabiduría estadística

STEPHEN M. STIGLER

Traducción de Miguel Nadal Palazón



Introducción

¿Qué es la estadística? Esta pregunta se planteó en fecha tan temprana como 1898 —refiriéndose a la Royal Statistical Society— y desde entonces se ha vuelto a plantear muchas veces. La persistencia de la pregunta y la variedad de respuestas que se le han dado a lo largo de los años son por sí mismas un fenómeno notable. Tomadas en conjunto, indican que la persistente perplejidad se debe a que la estadística no es una materia única. La estadística ha cambiado radicalmente desde sus primeros días hasta la actualidad, yendo de ser una profesión que reivindicaba una objetividad tan extrema que los estadísticos sólo reunirían datos —sin analizarlos—, hasta ser una profesión que busca asociarse con los científicos en todas las etapas de la investigación, desde la planeación hasta el análisis. Igualmente, la estadística presenta diferentes rostros a las diferentes ciencias: en algunas aplicaciones, aceptamos los modelos científicos como si provinieran de la teoría matemática; en otras, construimos un modelo que pueda adquirir luego un estatus tan sólido como cualquier construcción newtoniana. En algunas situaciones, somos planificadores activos y analistas pasivos; en otras, somos lo opuesto. Con tantas caras, y con las consiguientes dificultades para mantener el equilibrio y evitar tropiezos, no debe sorprendernos que la pregunta sobre qué es la estadística haya surgido una y otra vez, siempre que se enfrenta un nuevo reto, sean las estadísticas económicas de la década de 1830, sean las cuestiones biológicas de la de 1930 o las preguntas imprecisamente planteadas sobre *big data* en los tiempos que corren.

Dada la gran variedad de preguntas, aproximaciones e interpretaciones estadísticas, ¿acaso no existe un núcleo duro en la ciencia de la estadística? Si nos dedicamos de manera central a trabajar en tantas ciencias diferentes, desde el estudio de las políticas públicas hasta la validación del descubrimiento del bosón de Higgs, y si a veces se nos considera como simple personal técnico, ¿realmente podemos asumirnos, en algún sentido razonable, como practi-

cantes de una disciplina unificada, incluso de una ciencia por mérito propio? Ésta es la cuestión que quiero atender en este libro. No intentaré decir qué es o qué no es la estadística: intentaré formular siete principios, siete pilares que en el pasado han sostenido nuestra disciplina de diferentes maneras y que prometen hacerlo también en el futuro. Trataré de demostrar que cada uno de ellos fue revolucionario cuando se presentó, y que cada uno se mantiene como un avance conceptual importante y profundo.

Mi título se hace eco del libro de memorias de T. E. Lawrence, mejor conocido como Lawrence de Arabia: *Los siete pilares de la sabiduría*, de 1926.¹ Su relevancia estriba en la propia fuente de Lawrence, el libro de Proverbios del Antiguo Testamento (9:1), que dice “La sabiduría edificó su casa, labró sus siete columnas.”[†] De acuerdo con este versículo, se construyó la casa de la sabiduría para dar la bienvenida a todos aquellos que buscaban el entendimiento; mi versión tendrá una meta adicional: articular el núcleo intelectual del razonamiento estadístico.

Al llamar a estos siete principios los “siete pilares de la sabiduría estadística” me apresuro a enfatizar que se trata de siete pilares de *soporte*: son las bases de la disciplina, no el edificio completo de la estadística. Los siete tienen orígenes antiguos, y la disciplina moderna ha construido su multifacética ciencia sobre esta estructura con gran ingenio y con un suministro constante de ideas nuevas, excitantes, siempre prometedoras. Pero, sin restarle nada a esta obra moderna, espero articular la unidad en el núcleo mismo de la estadística, tanto a lo largo del tiempo cuanto entre sus áreas de aplicación.

Al primer pilar lo llamaré Agregación, aunque bien podría recibir el nombre decimonónico de “combinación de observaciones”, o

[†] La tradición religiosa ha consagrado en español la traducción “siete columnas”, a diferencia de las traducciones en lengua inglesa que se decantaron por *seven pillars*, “siete pilares”. Aunque para la cita de Proverbios 9:1 se ha optado por la versión Reina-Valera (que, además, es la más cercana a la Biblia del rey Jacobo seguida por Stigler), se mantendrá “pilares” en el resto del texto, en consonancia con el título, que a su vez está inspirado en el de Lawrence. [N. del t.]

incluso si se reduce al ejemplo más simple, el de “calcular una media”. Dichos nombres sencillos son engañosos pues hago referencia a una idea que ahora es vieja pero que fue auténticamente revolucionaria en tiempos anteriores, y lo es todavía, cada vez que alcanza una nueva área de aplicación. ¿De qué manera es revolucionaria? Lo es porque estipula que, dada una cantidad de observaciones, en verdad se puede obtener información ¡si se desecha información! Al calcular una simple media aritmética, descartamos la individualidad de las medidas, subsumiéndolas en otra que es un resumen. Hoy en día ello puede resultar natural en la astronomía con las mediciones sucesivas de, digamos, la posición de una estrella, pero en el siglo XVII se habría necesitado hacer la vista gorda con que la observación francesa la había hecho un observador propenso a la bebida y con que la observación rusa se había hecho con un instrumento viejo, mientras que la observación inglesa era de un buen amigo que jamás te había decepcionado. Los detalles de las observaciones individuales debían ser, efectivamente, borrados para revelar un indicador mejor que lo que cualquier observación individual podría ofrecer por sí misma.

El primer uso claramente documentado de la media aritmética tuvo lugar en 1635; otras formas de resumen estadístico tienen historias mucho más largas, que se remontan hasta Mesopotamia y casi hasta el amanecer de la escritura. Por supuesto, los ejemplos relevantes recientes de este primer pilar son más complicados. El método de los mínimos cuadrados y sus primos y descendientes son todos promedios ponderados de datos en los que se diluye la identidad de los individuos, salvo por su carácter de covariables. Y dispositivos tales como los estimadores de tipo núcleo de densidad y varios suavizadores modernos también son promedios.

El segundo pilar es la Información, más específicamente la Medición de la Información, y también tiene una larga e interesante historia. La pregunta de cuándo tenemos suficiente evidencia para convencernos de que funciona un tratamiento médico se remonta hasta los griegos. El estudio matemático de la tasa de acumulación de la información es mucho más reciente. A comienzos del siglo XVIII, se descubrió que en muchas situaciones la cantidad de información en un conjunto de datos era proporcional solamente a la

raíz cuadrada del número de observaciones n , y no al número n en sí mismo. Esto, también, resultó revolucionario: ¿se imagina tratar de convencer a un astrónomo de que si desea duplicar la precisión de una investigación necesita cuadruplicar el número de observaciones, o de que las segundas 20 observaciones no son ni de cerca tan informativas como las primeras 20, a pesar de que sean igual de precisas? A esto se ha dado en llamar la regla de la raíz de n ; se requirió de algunas conjeturas difíciles, y se necesitó hacer modificaciones en muchas situaciones complicadas. En cualquier caso, la idea de que la información en los datos podía ser medida, de que la precisión se relacionaba con la cantidad de datos de modo que podía ser articulada de manera precisa en algunas situaciones, ya estaba claramente establecida hacia 1900.

Con el nombre que le doy al tercer pilar, Verosimilitud, me refiero a la calibración de las inferencias usando la probabilidad. La forma más simple de ello está en las pruebas de significancia y en el habitual p -value o valor p , pero, tal como indica el nombre “verosimilitud”, hay abundantes métodos asociados, muchos de ellos relacionados con las familias paramétricas o con la inferencia, sea fisheriana o bayesiana. Aplicar pruebas de una forma o de otra se remonta mil años o más, pero algunas de las pruebas más tempranas que recurren a la probabilidad tuvieron lugar en los albores del siglo XVIII. Hay múltiples ejemplos en ese siglo y en el anterior, pero el tratamiento sistemático sólo llegó con los trabajos, ya en el siglo XX, de Ronald A. Fisher, por un lado, y de Jerzy Neyman y Egon S. Pearson, por otro, cuando se comenzó seriamente a elaborar una teoría completa de la verosimilitud. El uso de la probabilidad para calibrar las inferencias puede ser más conocido en las pruebas estadísticas, pero tiene lugar cada vez que un número está unido a una inferencia, sea un intervalo de confianza o una probabilidad *a posteriori* bayesiana. En efecto, el teorema de Thomas Bayes fue publicado hace unos 250 años justo con ese propósito.

Tomo prestado el nombre que le doy al cuarto pilar, Intercomparación, de un viejo artículo de Francis Galton. Representa lo que alguna vez fue una idea radical y ahora es un lugar común: las comparaciones estadísticas no necesitan realizarse respecto de un estándar externo, sino que frecuentemente se pueden llevar a cabo

dentro de los propios datos. Los ejemplos más comunes son las pruebas t de Student y las pruebas de análisis de varianza. En los diseños complejos, la partición de la variación puede ser una operación compleja y requerir la separación de bloques, las parcelas subdivididas o la evaluación de diseños jerárquicos basándose por completo en los datos disponibles. La idea es bastante radical, y la capacidad de ignorar estándares científicos externos puede llevar a abusos en las manos equivocadas, como ocurre con la mayoría de las herramientas poderosas. El *bootstrap* se puede entender como una versión moderna de la intercomparación, pero con supuestos más débiles.

Llamo al quinto pilar Regresión siguiendo el descubrimiento de Galton de 1885, explicado en términos de la distribución normal bivariada. Galton llegó a esto al intentar concebir un marco matemático para la teoría de la selección natural de Charles Darwin, que superara lo que a Galton le parecía una contradicción intrínseca de la teoría: la selección requería aumentar la variabilidad, lo que contradice la apariencia de estabilidad poblacional necesaria para la definición de las especies.

El fenómeno de la regresión se puede explicar brevemente: si se tienen dos medidas que no están perfectamente correlacionadas y se elige una como valor extremo respecto de la media, se espera que la otra sea menos extrema, medida en unidades de desviación estándar. Los padres altos en promedio producen hijos algo más bajos que ellos mismos; los hijos altos en promedio tienen padres algo más bajos que ellos mismos. Pero aquí está involucrado algo más que una simple paradoja: la idea verdaderamente novedosa fue que la pregunta daba respuestas radicalmente diferentes dependiendo de la forma en que se planteara. Dicho trabajo de hecho introdujo el análisis multivariado moderno, y las herramientas necesarias para cualquier teoría de la inferencia. Antes de la introducción de este aparato de distribuciones condicionadas, un teorema de Bayes auténticamente general no era viable. De ese modo, este pilar resulta medular para la inferencia bayesiana, lo mismo que para la causal.

El sexto pilar es el Diseño, pero entendido de manera más amplia que en expresiones como “diseño experimental”: se trata de un

ideal que puede disciplinar nuestro razonamiento incluso en entornos observacionales. Algunos elementos de diseño son extremadamente viejos. El Antiguo Testamento y la medicina arábica temprana ofrecen ejemplos. A partir de finales del siglo XIX surgió una nueva forma de entender el tema, conforme Charles S. Peirce y luego Fisher descubrían el extraordinario papel que la aleatorización podía tener en la inferencia. Al reconocer las ganancias que se podrían obtener de un acercamiento combinatorio con aleatorización rigurosa, Fisher llevó este asunto a nuevos niveles cuando introdujo cambios radicales en la experimentación que contradijeron siglos de concepciones y prácticas experimentales. En pruebas de campo multifactoriales, los diseños de Fisher no sólo permitieron la separación de los efectos y la estimación de las interacciones: el hecho mismo de la aleatorización hizo posibles inferencias válidas que no requerían que se supusiera la normalidad o la homogeneidad del material.

Llamo Residuo al séptimo y último pilar. Se podría sospechar que se trata de una evasión, si entendemos *residuo* como “todo lo demás”. Pero tengo algo más específico en mente. La noción de los fenómenos residuales se volvió común en los libros de lógica a partir de la década de 1830. Como lo planteó un autor: los fenómenos complicados pueden simplificarse al reducir el efecto de las causas conocidas y dejando un *fenómeno residual* por explicar; “Es principalmente mediante este proceso [que] la ciencia progresa.”² La idea, por tanto, es clásica en sus líneas generales, pero su uso en estadística tomó una nueva forma que acentúa y disciplina radicalmente el método al incorporar familias estructuradas de modelos y al emplear el cálculo de probabilidades y la lógica estadística para escoger entre ellos. Las expresiones más comunes en estadística son nuestros diagnósticos de modelos (gráficas de residuos), pero más importante aún es la manera en que exploramos espacios de muchas dimensiones al ajustar y comparar métodos anidados. Cada prueba de significancia de un coeficiente de regresión es un ejemplo, así como lo es cada exploración de una serie de tiempo.

Corriendo el grave riesgo de sobresimplificar, podría resumir y replantear estos siete pilares para representar la utilidad de siete ideas estadísticas básicas:

1. El valor de la reducción dirigida o la compresión de datos.
2. El decreciente valor de un creciente número de datos.
3. Cómo poner una varita de medir probabilística a lo que hacemos.
4. Cómo usar la variación interna en los datos para ayudar en el punto anterior.
5. Hacer preguntas desde distintas perspectivas puede conducir a respuestas reveladoramente diferentes.
6. El papel esencial de la planeación de las observaciones.
7. Cómo todas esas ideas pueden usarse para explorar y comparar explicaciones rivales en la ciencia.

Pero estos replanteamientos simplones no expresan cuán revolucionarias fueron estas ideas al ser descubiertas tanto en el pasado como en el presente. En todos los casos han hecho a un lado o anulado creencias matemáticas o científicas firmemente sostenidas, desde desechar la individualidad de los datos hasta reducir la ponderación de datos nuevos e igualmente valiosos, o hasta superar las objeciones hacia cualquier uso de la probabilidad para medir la incertidumbre fuera de los juegos de azar. ¿Y cómo puede la variabilidad interna de nuestros datos medir la incertidumbre del mundo que los produjo? El análisis multivariado de Galton les reveló a los científicos que su confianza en las reglas de proporcionalidad de la época de Euclides no funcionaba en un mundo científico donde existe la variación en los datos, derribando así tres mil años de tradición matemática. Los diseños de Fisher contradecían directamente lo que, por siglos, habían creído los científicos experimentales y los lógicos; sus métodos para comparar modelos eran completamente nuevos para la ciencia experimental y se requirió un cambio generacional para que fueran aceptados.

Como evidencia de lo revolucionarias e influyentes que fueron estas ideas, considérense las fuertes resistencias que aún atraen, y que frecuentemente atacan los mismos aspectos que presento como características valiosas. Me refiero a:

- ♦ las quejas acerca del desdén a los individuos, el trato a las personas como meras estadísticas;

- ♦ las afirmaciones implícitas de que el *big data* puede dar respuestas gracias sólo al tamaño;
- ♦ las denuncias de que las pruebas de significancia niegan la ciencia en cuestión, y
- ♦ las críticas a los análisis de regresión por desatender aspectos importantes del problema.

Estas cuestiones son problemáticas, pues las acusaciones pueden llegar a ser correctas y estar bien dirigidas en el caso que las motiva, pero con frecuencia están dirigidas al método, no a la manera en que se ha usado en tal o cual situación. Edwin B. Wilson hizo un buen comentario a este respecto en 1927: “Es en gran medida debido a la falta de conocimiento sobre qué es la estadística que el individuo no entrenado en su uso se confía a una herramienta tan peligrosa como cualquier otra que podría elegir entre todo el arsenal de la metodología científica.”³

Los siete pilares que describiré, y cuya historia voy a esbozar, son herramientas finas que requieren manos sabias y bien entrenadas para su uso efectivo. Dichas ideas no son parte de las matemáticas, ni son parte de la ciencia de la computación. Son fundamentalmente de la estadística, por lo que debo confesar que, aunque comencé por negar explícitamente que mi objetivo fuera explicar qué es la estadística, al término de este libro tal vez haya hecho eso.

Vuelvo brevemente sobre un cabo suelto: ¿qué significa exactamente el pasaje de Proverbios 9:1? Es una aseveración extraña: “La sabiduría edificó su casa, labró sus siete columnas.” ¿Por qué una casa necesitaría siete pilares, una estructura desconocida, según parece, tanto en tiempos pasados como hoy en día? Investigaciones recientes han mostrado, de manera convincente en mi opinión, que los estudiosos, incluyendo los traductores que produjeron la Biblia de Ginebra y la versión del rey Jacobo, no conocían bien la mitología sumeria temprana, y tradujeron equivocadamente este pasaje en el siglo XVI. No se hacía referencia a ninguna estructura material; más bien se refería a los siete grandes reinos de Mesopotamia antes del diluvio, siete reinos en siete ciudades fundadas sobre principios formulados por siete sabios que aconseja-

ban a los reyes. La casa de la sabiduría se basaba en los principios de estos siete eruditos. Un estudioso más reciente ha propuesto esta traducción alternativa: “La sabiduría ha construido su casa, Los siete han puesto sus cimientos.”⁴

Justamente así, los siete pilares que planteo son el fruto de los esfuerzos de muchos más que siete sabios, incluyendo algunos cuyos nombres se han perdido para la historia, y en estas páginas conoceremos a un buen número de ellos.