

## COMENTARI DE LLIBRES

**C.R. Rao.**

### **STATISTICS AND TRUTH.**

Council of Scientific and Industrial Research, New Delhi, 1989, 143 páginas, 15 tablas, 6 figuras.

En un intento de situar el libro "Statistics and Truth" de C. Radhakrishna Rao, doy una clasificación personal y desde luego no exhaustiva, de los libros de Estadística.

- A) Libros escritos por y para especialistas, que no se entienden salvo si se domina el análisis funcional, la geometría diferencial o la teoría de grupos.
- B) Libros que conectan con la realidad de los datos, escritos en un lenguaje claro y asequible, de gran ayuda tanto en la función docente como en la actividad investigadora.
- C) Libros escritos con la propiedad de contener errores conceptuales que no siguen las leyes del azar sino las leyes de la ignorancia de sus autores.

Los libros de la clase A son útiles para los dodecafónicos de la Estadística, sólo preocupados de armonizar con la matemática abstracta creando, es un decir, líneas melódicas que algunos entienden pero pocos sienten. Peores son los de la clase C, a cuyos autores hay que perdonar porque no saben lo que se escriben, libros sumamente útiles para una minoría recalcitrante de usuarios que ofician imposibles casamientos entre datos reales y métodos clásicos, cuyos frutos son resultados híbridos que engrosan la amplia nómina de la investigación estéril. Ni los A ni los C se salvarían, como el Tirant lo Blanc, de la hoguera del olvido, si se nos permite imaginar que el ingenioso hidalgo fuese estadístico. Sí, en cambio, merecen ser salvados y alabados los del grupo B mencionado, como el reciente libro de C.R. Rao.

"Statistics and Truth" es un ejemplo paradigmático. Bien escrito, nivel asequible pero no elemental, algo informal, es extraordinariamente rico en ideas, reflejando la visión que la dilatada experiencia proporciona al autor, quizás el mejor estadístico en la actualidad.

“Statistics and Truth” está escrito en tono polémico, amable pero incisivo, debido a que recoge tres conferencias del autor dictadas en memoria de Srinivasa Ramanujan (1887-1920) un raro fenómeno, como bien dice, pues era un genial matemático indio que descubrió, al principio de forma autodidacta, increíbles fórmulas de la teoría de números. La admiración de Rao por Ramanujan está presente en toda la obra.

El primer capítulo-conferencia, titulado “Uncertainty, randomness and creation of new knowledge”, es una interesante meditación filosófica sobre la aleatoriedad e incertidumbre. La duda de que el azar es una expresión de nuestra ignorancia queda disipada. El azar es inherente a todos los fenómenos naturales y la única manera de realizar predicciones óptimas es conocer sus estructuras y formular reglas apropiadas de acción. Tal concepción lleva al autor a añadir al dicho: “Las Ciencias son, en abstracto, Matemáticas”, la siguiente frase: “Todos los métodos de adquisición de conocimientos recurren esencialmente a la Estadística”. Luego la verdad se basa muy a menudo en la Estadística. Por ejemplo, la interpretación estadística de la segunda ley de la termodinámica, dada por Boltzman, así como el uso de la probabilidad dados por Quetelet y Mendel, permiten justificar el paso del “determinismo al orden en desorden”, es decir, cómo la incertidumbre a nivel individual nos lleva, en media, a una cierta estabilidad por obra y gracia de la ley de los grandes números, que Rao ilustra haciendo énfasis en la teoría de Bose-Einstein. Una amplia discusión sobre la casi inexplicable influencia de la aleatoriedad en la creatividad cierra el capítulo. Se añade en forma de apéndice una discusión sobre caos y azar, creatividad, azar y necesidad, niveles de incertidumbre y respuestas aleatorizadas a preguntas sensibles.

El segundo capítulo-conferencia, titulado “Principles and Strategies of Data Analysis, Part 1: Cross-Examination of Data”, es un verdadero compendio de Estadística. Empieza con una ingeniosa observación: los estilos de tratar los datos cambian, pero los datos permanecen, para continuar diciendo que la Estadística no es aún lo bastante añeja para ser considerada una disciplina estable. Sin embargo, añade, a pesar de las controversias, la Estadística es una disciplina en franca expansión, en buena parte gracias a la informática. Con abundantes referencias a nuestros primeros padres (gracias a los cuales los estadísticos existimos y nos criticamos entre sí, pero menos de lo que nos critican los demás), es decir, K. Pearson, Fisher, Mahalanobis, Hotelling, Bose, Neyman, Roy, Wilks, Wald, a quienes, en su mayoría, Rao conoció y trató, distingue entre el Análisis Descriptivo de Datos y el Análisis Inferencial de Datos, enunciando el principal principio del Análisis de Datos en forma de ecuación:

Data Analysis = Answering Specific Questions  
+ Providing Information for New Lines of Research.

Rao hace énfasis en la *cross-validation* de los datos, revisando todas aquellas virtudes y defectos del recolector, comentando interesantes aspectos a tener en cuenta y denunciando, con amabilidad pero con precisión, las trampas de ciertas investigaciones (desde el poco conocido Lazzarini en su sospechosa estimación de  $\pi$  hasta el propio Mendel, que publicó datos tan bien ajustados a sus teorías que hacen pensar en alguna censura eclesiástica). Atención especial es dada a los outliers, el uso de gráficos (la forma más rápida de engañar con estadísticas) y los modelos de regresión, tan ampliamente usados como mal utilizados (especialmente por los usuarios de los libros *C* antes comentados, capaces de plantear una regresión lineal con sólo dos puntos sin inmutarse). Son muy oportunas las observaciones en clave de humor que Rao recoge de Ch. Babbage, el pionero de la informática, acerca de las actitudes del científico que nunca sería armado caballero: (i) *Trimming*, (ii) *Cooking*, (iii) *Forging*, que ilustran irónicamente el arte de arreglar o desechar las observaciones que no interesan e inventar las que hacen falta, dada su resistencia a ser observadas con aparente antipatía hacia el investigador que tanto las necesita.

El capítulo termina con algunos comentarios sobre la especificación, es decir, la elección del verdadero modelo estocástico para trabajar con los datos observados. Rao nos hace ver que la elección del modelo debe estar equilibrada con el error de estimación, y que a veces el uso del verdadero modelo puede llevar a una pérdida de eficiencia. Recomienda la técnica de la *cross-validation* para validar el modelo y finaliza con un resumen de los aspectos más relevantes del Análisis Inferencial de Datos.

El tercer capítulo corresponde también a la segunda conferencia y se titula: "Principles and Strategies of Data Analysis, Part 2: Weighted and Clouded Distributions". Contiene la parte más técnica del libro. De nuevo plantea el problema de la especificación. Una especificación incorrecta lleva a una inferencia también incorrecta, lo que se ha venido en llamar el error de tercera clase. Sin embargo, argumenta Rao, el problema no es simple, pues ciertos sucesos pueden no ser observables o la probabilidad de ser observados variar según las circunstancias. Entonces las reglas del azar deben ser modeladas apropiadamente para el análisis estadístico. En esta línea se presenta la truncación de una distribución y las distribuciones con pesos (donde una observación  $x$  tiene una densidad de presentarse y una densidad  $w$  de ser registrada), que son ilustradas con ejemplos amenos. Una interesante aplicación hace referencia al cociente  $B/(B+S)$ , que resulta de preguntar a cada varón de una clase (o en cualquier reunión) el número

de hermanos (incluyéndose a sí mismo) y el número de hermanas y anotando por B y S el total de hermanos y hermanas respectivamente. Entonces, a lo largo de varias tablas y dos teoremas empíricos, el autor prueba que B es mucho mayor que S y que el cociente  $B/(B+S)$  es del orden de 0.6 o incluso  $2/3$ . Esta predicción, válida incluso si el número de varones en una clase es pequeño, (en el caso mujeres basta tomar  $S/(B+S)$ ), constituye un buen ejemplo didáctico para el profesor de Estadística. Otras aplicaciones, tanto o más interesantes, hacen referencia a la estimación del tiempo medio de estancia de turistas en un país, recuento de animales por muestreo de áreas cuadradas, obtención de medidas sobre restos óseos y probablemente rotos y medidas de temperaturas del mar mediante satélite que son afectadas por la presencia de nubes, dando lugar a las llamadas distribuciones nubladas. Se trata, en definitiva, de un capítulo que plantea situaciones y propone soluciones que difícilmente se encuentran en otros libros, pero que sí aparecen en las aplicaciones prácticas, lo que constituye un perfecto complemento de la literatura estadística usual.

La tercera conferencia es el cuarto capítulo, cuyo título "Statistics: an Inevitable Instrument in Research of Truth" es bastante explícito. Parte menos técnica, algo retórica, va dirigida al gran público y demuestra como la Estadística juega un importante papel en la búsqueda de la verdad: en la ciencia, en el arte, en la política, en la industria, en los negocios, en la medicina, en las leyes, en fin, en toda la sociedad, demostrando la ubicuidad de la Estadística, que ha dado, según R.A. Fisher, un caracter especial al siglo XX.

Los ejemplos que nos ofrece Rao de la omnipotencia de la Estadística son tan numerosos como interesantes: cómo decidir si un nuevo poema puede ser asignado a Shakespeare; cómo decidir la autoría de un trabajo anónimo frente a diversos posibles autores; el problema de la filiación de manuscritos; el problema de la filogenia de los lenguajes indo-europeos a partir de un lenguaje ancestral común hablado hace unos 4.500 años; cómo se descubrió, aplicando mera estadística elemental, que las anguilas de los fiordos noruegos provenían de un criadero común en el océano que fué descubierto más tarde; una explicación de las proporciones de las formas levógiras y dextrógiras de enrollarse ciertas plantas; cómo explicar el ritmo circadiano; el problema de la disputa de la paternidad en los tribunales, etcétera. La recopilación, cuajada de anécdotas y citas históricas, es muy ámena.

La obra finaliza con un Apéndice de catorce páginas que viene a ser tanto una declaración de principios como una revisión de sus raíces y fundamentos. Tratando de definir qué es la Estadística, aquellas frases maliciosas antaño utilizadas ("Mentiras, condenadas mentiras y estadísticas", "Conozco la respuesta, dadme una estadística y os la justificaré") van siendo sustituidas por otras más ajustadas a conectar con la sociedad ("Según las estadísticas, los hombres que

permanecen solteros viven diez años menos”, “Según una encuesta, tomar aspirina reduce el riesgo de ataque al corazón”), con lo cual la estadística en sentido vulgar viene a apoyar o dar apariencia de validez (verdad) a las afirmaciones y frases corrientes. Remontándose al antiguo Egipto, la Estadística tiene una larga historia, que influye en la revolución industrial del siglo pasado, desembocando en la creación en 1885 del International Statistical Institute, la más alta institución dedicada a estadística oficial cuyos miembros deben ser electos y que se reúne cada dos años. En el intento de discernir si la Estadística es Ciencia, Tecnología o Arte, Rao concluye que es las tres cosas a la vez. Es, efectivamente, también Arte, pues diferentes estadísticos pueden llegar a diferentes conclusiones trabajando con los mismos datos. La libertad del Arte es también patrimonio de la Estadística. Como punto final, unas concisas consideraciones sobre los razonamientos deductivo e inductivo, este segundo base de la Estadística. Rao aprovecha para lanzar unas salvas de artillería (que en realidad son obuses) contra la llamada inferencia deductiva (léase enfoque bayesiano), con argumentos bastante convincentes y citando una romántica anécdota que prueba cuán desastroso puede llegar a ser el uso de la información a priori.

La lectura de “Statistics and Truth” es amena, rigurosa, ingeniosa, actual y muy recomendable para cualquier científico y profesional deseoso de conocer el papel de la Estadística en la búsqueda de la Verdad, sin renunciar a un cierto sentido del humor. (Anécdota en la página 44: preguntada la secretaria del ministro de la salud que explique sobre la afirmación de que 6.5 de cada 1000 habitantes podrían morir de sarampión, ella responde que 6.5 significa que 6 fallecen y 5 están a punto de fallecer). Sin embargo, el libro es algo difícil de encontrar. Esperemos que pronto aparezca una nueva edición, incluso traducción, más fácil de conseguir.

C.M. Cuadras.