

VERSIÓN INGLESA

El arte de la estadística

Maite Aranés Maza

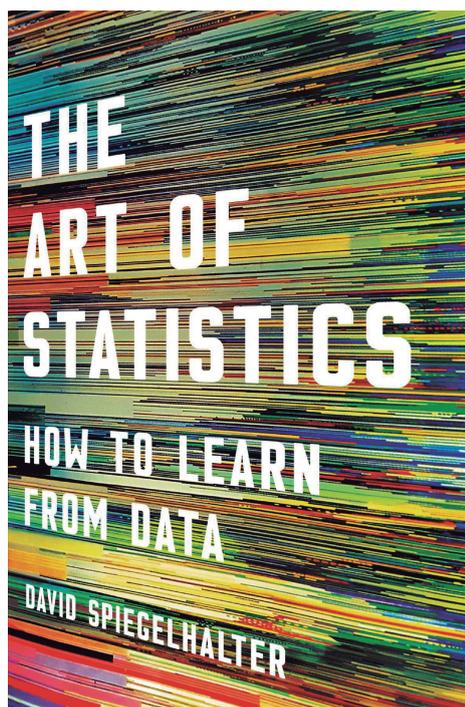
SUMA núm. 95
pp. 125-130

Artículo solicitado por *Suma* en julio de 2020 y aceptado en septiembre de 2020

En estos últimos tiempos hemos sido bombardeados con números y estadísticas desde los medios aún más de lo habitual. Esto me ha servido de motivación para leer mi primer libro de divulgación sobre estadística, al que me he acercado con igual nivel de expectación y aprensión. Las razones para escoger este volumen en particular han sido la fecha reciente de publicación y las credenciales del autor.

Rompiendo la tónica de las anteriores reseñas, el autor de *The Art of Statistics* no es (ni ha sido) profesor de secundaria. David Spiegelhalter es un reconocido estadístico británico, presidente de la Royal Statistical Society entre 2017 y 2019, y con gran dedicación en la labor de la divulgación científica.

En la introducción descubrimos que la intención del autor es introducir las ideas fundamentales de la estadística utilizando problemas reales, evitando detalles técnicos y concentrándose en explicarnos los con-



Pelican, 448 páginas (2020)

ceptos subyacentes. En principio esto me parece un objetivo muy loable, puesto que claramente quiere alcanzar a un público lo más amplio posible. Entre ese público Spiegelhalter incluye específicamente a los estudiantes de estadística. Esto enlaza además con sus comentarios sobre la enseñanza de la estadística. En su opinión, en el pasado se ha caído a menudo en la trampa de limitarse al uso de herramientas y explicación de teorías matemáticas:

Generations of students have suffered through dry statistics courses based on learning a set of techniques to be applied in different situations, with more regard to mathematical theory than understanding both why the formulae are being used, and the challenges that arise when trying to use data to answer questions.

Sin embargo, por más que sus intenciones son admirables, al leer la introducción una no puede evitar sentir cierta inquietud por la forma en que enfatiza que el libro contiene solo un número muy pequeño de ecuaciones inofensivas, y que no vamos a necesitar habilidades matemáticas. ¿Será capaz de explicarse con claridad suficiente contando con estas restricciones? Lo cierto es que me resulta difícil juzgar, porque no soy precisamente una experta en estadística. En los temas que me resultan más familiares, me ha parecido que ha hecho un gran trabajo.

Por otra parte, Spiegelhalter es un maestro a la hora de captar nuestro interés utilizando situaciones y ejemplos de datos reales. De hecho, su introducción no comienza con el propósito del libro, que viene al final del capítulo, sino con un ejemplo muy llamativo del uso de la estadística: un análisis del perfil y hora de la muerte de las víctimas de Harold Shipman, el más prolífico asesino en serie de Gran Bretaña. Spiegelhalter formó parte del panel de expertos consultado para determinar el número de víctimas de este médico de cabecera (un escalofriante número, por encima de 200), lo que explica la inclusión de este ejemplo. A lo largo del libro nos encontramos una gran variedad de preguntas cuyas respuestas o análisis requieren el uso de herramientas estadísticas. Además de numerosas cuestiones relacionadas con la medicina, Spiegelhalter incluye temas que van de la exis-

tencia del bosón de Higgs hasta el tamaño de las orejas de los ancianos.

En los dos primeros capítulos, «Getting Things in Proportion: Categorical Data and Percentages» y «Summarising and Communicating Numbers. Lots of Numbers», se dedica a explorar la representación de datos, tanto a través de gráficos como de tablas y medidas numéricas. Por supuesto buena parte de las ideas expuestas nos resultarán muy familiares, pero por mi parte he encontrado algunos aspectos desconocidos o con un enfoque algo distinto. Por ejemplo, uno de los puntos más destacados del primer capítulo es la comunicación de riesgo a través de porcentajes. Spiegelhalter recomienda que para la difusión al público se utilicen riesgos absolutos en lugar de relativos. Para ilustrar el problema recurre al incremento en el riesgo de desarrollar cáncer de colon como resultado de consumir un bocata de beicon diario (o 50 g de carne procesada). En términos absolutos el riesgo pasa de ser de un 6 % a un 7%, y en términos relativos el riesgo aumenta un 18%. La segunda cifra puede resultar alarmista, y nos hacemos una idea más clara con la primera comparación. En esta discusión Spiegelhalter también habla de la «odds ratio» (me disculpo porque no conocía este término, y no estoy segura de cuál es la traducción correcta) (figura 1). Esta proporción se utiliza habitualmente en estudios epidemiológicos y alguna vez los medios de comunicación la han reproducido de forma errónea como riesgo relativo.

Method	Non-bacon eaters	Daily bacon eaters
Event rate	6%	7%
Expected frequency	6 out of 100	7 out of 100
	1 in 16	1 in 14
Odds	6/94	7/93
Comparative measures		
Absolute risk difference	1%, or 1 out of 100	
Relative risk	1.18, or an 18% increase	
'Number Needed to Treat'	100	
Odds ratio	$(7/93) / (6/94) = 1.18$	

Figura 1. Cuando el evento estudiado es poco común, la «odds ratio» es muy similar al riesgo relativo, como sucede en este ejemplo

El tercer capítulo ofrece una discusión muy detallada sobre las ideas de muestra y población, y sobre los problemas que podemos encontrarnos en la recolección de datos. Explica el concepto de razonamiento inductivo, que consiste en obtener conclusiones generales (sobre una población) a partir de datos particulares (obtenidos por ejemplo de una muestra). Utilizando ejemplos para ilustrar esta idea introduce conceptos más complejos como distribución de una población y distribución de probabilidad, y la distinción entre estadístico muestral y parámetros estadísticos.

En el capítulo siguiente, «What Causes What?» Spiegelhalter hace un trabajo estupendo hablando de asociaciones entre variables, y dejando bien claro que correlación no implica causalidad. Aprovecha también para describir las características que debería tener un ensayo clínico bien diseñado, utilizando como ejemplo estudios sobre la efectividad de las estatinas. Esta discusión le lleva de forma natural al tema del capítulo 5, la regresión como modelo estadístico. En este punto yo echo de menos alguna ecuación. Cuando Spiegelhalter introduce la recta de regresión, necesita varias líneas para explicarnos de forma aproximada la relación entre la pendiente de la recta y el coeficiente de correlación. Como consecuencia en algunos momentos el texto se hace algo farragoso, y en otros se dificulta la interpretación del ejemplo que está utilizando. Debo puntualizar que la ecuación de la recta se puede encontrar en el glosario de «términos técnicos» que incluye al final del libro, pero personalmente hubiese apreciado el tenerla delante durante la discusión del ejemplo.

Un aspecto que no he mencionado todavía es la inclusión de numerosos detalles sobre la historia de la estadística que Spiegelhalter va contando de forma muy amena durante todo el texto. Por ejemplo en el capítulo 5 utiliza el mismo conjunto de datos sobre alturas de padres e hijos con el que Francis Galton introdujo la idea de regresión a la media.

El capítulo 6, «Algorithms, Analytics and Prediction», trata de eso, algoritmos y predicciones. Dada mi ignorancia en el tema no puedo juzgar si se ha dejado algún aspecto esencial, o si ha sido poco preciso. En

cualquier caso resulta una lectura muy interesante, donde Spiegelhalter introduce ideas como aprendizaje automatizado (*machine learning*) y evaluación de la eficiencia de un algoritmo. Como ha hecho en los capítulos anteriores, los distintos conceptos y herramientas se introducen a través de un ejemplo concreto, en este caso, el diseño de un algoritmo que predice qué pasajeros sobrevivieron el naufragio del Titanic (figura 2). En las últimas secciones nos comenta las principales preocupaciones y problemas que podemos encontrarnos al utilizar un algoritmo.

Uno de los temas centrales del libro, al que Spiegelhalter alude una y otra vez, es el problema de variabilidad en los datos. Desde la primera mención de variabilidad en la introducción, vamos encontrando la misma cuestión en cada capítulo. Resulta natural pues, que sea el tema central en los siguientes capítulos, donde Spiegelhalter analiza con detalle las ideas de intervalo de confianza, márgenes de error y test de hipótesis. Comienza con una introducción muy práctica en el capítulo 7, «How Sure Can We Be About What Is Going On? Estimates and Intervals». Al final del primer ejemplo ya nos deja claro que

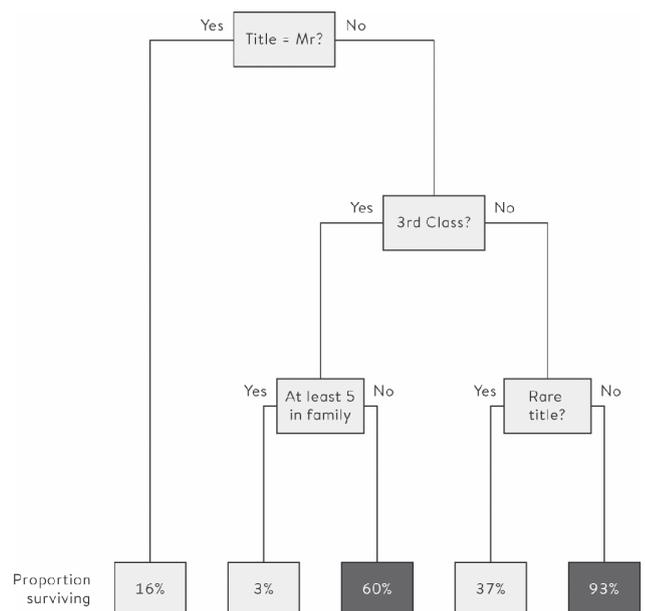


Figura 2. Árbol de clasificación diseñado a partir de datos sobre los supervivientes del Titanic. Un árbol de clasificación es un tipo de algoritmo de predicción (según Spiegelhalter, posiblemente el tipo más sencillo)

vamos a empezar a ver un aspecto crucial de la estadística, y nos advierte que quizá tenga que explicarnos ideas bastante complejas.

Acknowledging uncertainty is important. Anyone can make an estimate, but being able to realistically assess its possible error is a crucial element of statistical science. Even though it does involve some challenging concepts.

Este capítulo a mí me parece excelente. Spiegelhalter es conciso y claro, y consigue además transmitir interés sobre el tema. No me pasa muy a menudo que la explicación de un método de remuestreo me produzca un momento de «¡ajá!», ni que la mención del teorema central del límite me haga decir «Oooohhh qué chulo» (figura 3).

Spiegelhalter continúa con una introducción a la probabilidad en el siguiente capítulo. Una buena parte del material nos va a resultar muy familiar, porque nos está contando la probabilidad que explicamos en la educación secundaria. Eso sí, con menos formulitas, que él evita sobre todo apoyándose en árboles de frecuencias para ejemplificar ideas. Es en las últimas secciones donde nos encontramos con ideas más sutiles: ¿qué es la probabilidad?, ¿por qué necesitamos probabilidad si tenemos todos los datos?

Llegamos ahora a la parte más «dura» del libro, los capítulos 9 y 10. Por supuesto, los lectores avezados en los entresijos de la ciencia estadística tal vez no encuentren gran diferencia con el resto del libro. Pero ya os he avisado de que yo no me encuentro entre estos afortunados. En estos capítulos Spiegelhalter requiere nuestra completa atención. Introduce ideas más sofisticadas y nos pide además que seamos capaces de seguirle cuando va haciendo referencias y comparaciones entre los nuevos conceptos y lo explicado anteriormente, y con respecto a ejemplos y casos ilustrados en otros capítulos.

Dicho esto, os resumo como siempre algunos de los aspectos de estos capítulos. En «Putting Probability and Statistics Together» comenzamos con la distribución binomial, y el mismo ejemplo sirve de base para introducir la esperanza, el error estándar, la ley

de grandes números y el teorema central del límite. Esto en la primera mitad del capítulo. La segunda mitad está dedicada a la definición y análisis del concepto de intervalo de confianza. Todo esto sin incluir ni una fórmula. El hecho de que Spiegelhalter consiga explicarlo todo de manera clara, y dándonos una visión integral de estas ideas, dice mucho, en mi opinión, del gran trabajo que ha realizado en este texto. Aunque en teoría el material del capítulo me resulta familiar, he descubierto muchos aspectos sobre los que nunca había reflexionado. Por ejemplo, el autor dedica un par de páginas a aclarar la diferencia entre la incertidumbre aleatoria (antes de tirar una moneda no sabes qué resultado obtendrás) y la incertidumbre epistémica (después de tirar la moneda, mientras no

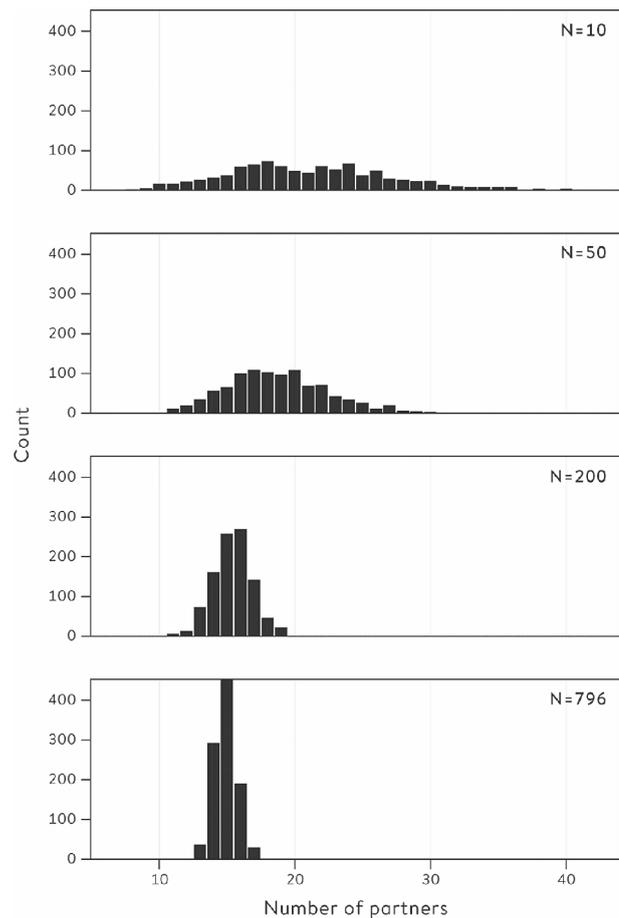


Figura 3. Distribución de las medias muestrales resultantes de 1000 remuestreos de una serie de datos, partiendo de muestras de tamaño 10, 50, 200 y 796 (el tamaño total de la muestra original). Se aprecia cómo la variabilidad de las medias obtenidas en los remuestreos disminuye cuando aumenta el tamaño de la muestra

la mires no sabes qué resultado has obtenido). Es en el segundo caso cuando recurrimos a la estadística.

So probability theory, which tells us what to expect in the future, is used to tell us what we can learn from what we have observed in the past. This is the (rather remarkable) basis for statistical inference.

Otro aspecto al que Spiegelhalter dedica bastante atención es el cálculo del margen de error de una estimación. Utilizando un ejemplo práctico ilustra las debilidades en limitarse a utilizar una fórmula estándar. Es preciso tener en cuenta también la fiabilidad de los datos, que ninguna fórmula va a corregir.

These insights should encourage us to have humility about the statistical methods we can bring to a single data source. If there are fundamental problems with the way the way the data have been collected then no amount of clever methods can eliminate these biases, and we have to use our background knowledge and experience to temper our conclusions.

El capítulo 10, «Answering Questions and Claiming Discoveries», es prácticamente el doble de largo que el resto. Esto se debe sobre todo a que el autor repasa muchos de los ejemplos y casos estudiados a lo largo del libro, analizando sus resultados con las nuevas herramientas introducidas en el capítulo anterior y en este. A este respecto, al principio del capítulo Spiegelhalter termina de sacar toda la artillería, y nos explica el proceso de contraste de hipótesis, el valor p y la noción de resultado estadísticamente significativo. Como viene siendo habitual, después de introducir las nuevas ideas a partir del estudio de varios casos concretos, Spiegelhalter se dedica a ilustrar los problemas y limitaciones de estas técnicas. Un ejemplo son los errores en la correcta interpretación del valor p y el concepto de falso positivo, ideas que entran en juego para entender el «problema de la repetición de pruebas».

Spiegelhalter nos explica que, por ejemplo, en el caso de un ensayo clínico, un valor p menor de 0,05 quiere decir que la probabilidad de que el medicamento que se esté estudiando no sea efectivo y obtengamos la conclusión de que es efectivo, es de un

5%. Es decir, la probabilidad de obtener un falso positivo es un 5%. El problema aparece si se repite el ensayo varias veces; aplicando las leyes de la probabilidad, si el ensayo se repite 10 veces la probabilidad de obtener un falso positivo en alguna de las repeticiones tomando $p < 0,05$ resulta ser aproximadamente un 40%. La situación es la misma cuando los investigadores dividen los datos en varios subconjuntos, realizan un contraste de hipótesis en cada uno y se fijan en el resultado más significativo. Aquí el autor nos da un ejemplo muy curioso, y no puedo resistir la tentación de compartirlo. En 2009 un equipo de investigadores realizaron un experimento con el objetivo de ilustrar el problema de repetición de pruebas y los falsos positivos. Se escaneó el cerebro de un sujeto mientras se le enseñaban una serie de fotos que ejemplificaban distintas emociones humanas, con la idea de identificar qué partes del cerebro mostraban una respuesta significativa, con $p < 0,001$. El caso es que el sujeto era un salmón de unos 2 kg de peso, que no estaba vivo en el momento de la prueba. Aún así como analizaron unas 8000 partes del cerebro del pescadito, identificaron 16 indicando una respuesta significativa a las fotografías... (figura 4).

Con esto llegamos a la que parece la parte más personal del libro, el capítulo 11, sobre inferencia bayesiana, «Learning from Experience the Bayesian Way». Spiegelhalter nos confiesa enseguida su preferencia por este tipo de inferencia estadística, y procede a introducir la idea de probabilidad condicionada, usando nuevamente árboles de frecuencias. A continuación entramos ya en materia, con la de-

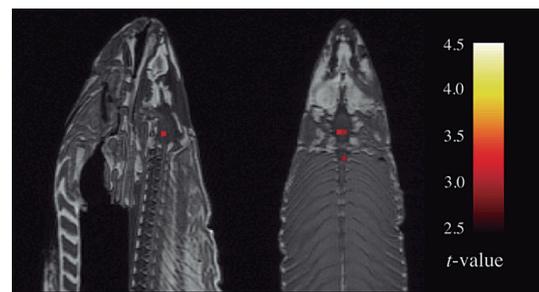


Figura 4. El resultado del experimento del salmón de Craig M. Bennett (se puede encontrar más información en el blog <prefrontal.org>)

finición de la proporción de verosimilitud (*likelihood ratio*) y su posible uso en el análisis forense, seguido de una explicación del proceso de inferencia bayesiana. Esta última parte me ha resultado difícil de seguir, y me he quedado solo con una idea general de los principios detrás de esta forma de hacer estadística. El capítulo concluye con una breve e interesante historia de la polémica entre la inferencia bayesiana y la frecuentista, que según el autor se ha suavizado, existiendo actualmente un enfoque más moderado en estas cuestiones dentro de la comunidad estadística.

The statistical community used to engage in lengthy vituperative arguments about the foundations of the subject, but now a guarded truce has been called and a more ecumenical approach is the norm, with methods chosen according to the practical context rather than their ideological credentials derived from Fisher, Neyman-Pearson or Bayes.

Los tres últimos capítulos son de lectura más ligera que los anteriores en cuanto a matemáticas se refiere. En el capítulo 12, «How things go wrong», se resumen y especifican los diversos problemas de la estadística que se han ido mencionando a lo largo de todo el libro. Spiegelhalter nos explica sin pelos en la lengua todos los puntos donde el asunto puede torcerse, desde los fallos en la investigación estadística hasta los problemas con la difusión de los resultados, tanto por parte de los medios como por las instituciones académicas. Por supuesto, una vez nos cuentan

los problemas, nos gusta ver soluciones. El autor lo entiende perfectamente y en el capítulo siguiente nos explica las medidas que pueden tomarse (muchas propuestas desde la comunidad científica) para corregir los problemas mencionados en el capítulo anterior. Aunque lo cierto es que no hay soluciones perfectas, como él mismo reconoce, por ejemplo en el caso de los medios de comunicación.

There is no simple way to influence the practice of journalism and the media, particularly in a time when the industry is being challenged due to competition from social media and unregulated online publications, and advertising revenues are dwindling, but statisticians have collaborated on reporting guidelines for media organisations and training programs for journalists and press officers.

El libro termina con un brevísimo capítulo de conclusión, que apenas llega a las dos páginas, seguido por un glosario de términos técnicos, una lista de referencias por capítulos y un índice alfabético.

Espero haber dado una idea de la amplitud del material que repasa el libro. Para apreciar lo bien que se explica Spiegelhalter, tendréis que leerlo vosotros. Por mi parte, comentaros que *The Art of Statistics* aparte de informarme, también ha conseguido cambiar mi percepción de la estadística. Ha pasado de ser «algo útil de lo que debería saber más» a «algo interesante de lo que quiero saber más». Y eso, la verdad, no me lo esperaba.

Maite Aranés Maza

IES Mar de Aragón, Caspe (Zaragoza)

<maite.aranes@gmail.com>