

PRÓLOGO

Estadística aplicada básica (versión en castellano de *The Basic Practice of Statistics*) es una introducción a la estadística para estudiantes universitarios que enfatiza el trabajo con datos y las ideas estadísticas. En este prólogo describo el contenido del libro con el objeto de ayudar a los profesores a juzgar si este libro es adecuado para sus alumnos.

Principios básicos

Hasta ahora, los cursos de introducción a la estadística se centraban en la probabilidad y en la inferencia. *Estadística aplicada básica* refleja una nueva corriente en la enseñanza de la estadística en la que el análisis de datos y el diseño de métodos para su obtención constituyen, junto con los métodos de inferencia basados en la probabilidad, los temas principales. Los estadísticos han llegado a un amplio consenso sobre los contenidos de cursos universitarios de introducción a la estadística. Tal como dice Richard Schaffer: “con relación al contenido de un curso de introducción a la estadística, los estadísticos están mucho más de acuerdo ahora que cuando yo empecé a trabajar”.¹ La figura 1 es un esquema de este consenso de acuerdo con la Sociedad Americana de Estadística (ASA) y la Asociación Americana de Matemáticas (MAA).²

Como miembro que fui del comité conjunto de la ASA/MAA, estoy de acuerdo con sus conclusiones: aunque el trabajo con datos puede suponer una ayuda, fomentar el aprendizaje activo es tarea del profesor. Por ello, las dos primeras recomendaciones son los principios directores de este texto. Si bien el libro es

¹D. S. Moore *et al.*, “New pedagogy and new content: the case of statistics”, *International Statistical Review*, 65, 1997, págs. 123-165. El comentario de Richard Scheaffer aparece en la página 156.

²George Cobb, “Teaching statistics” en L. A. Steen (ed.), *Heeding the Call for Change: Suggestions for Curricular Action*, Mathematical Association of America, Washington, D.C., 1990, págs. 3-43.

elemental con relación al nivel de matemáticas exigido y a los procedimientos estadísticos que se presentan, aspira a facilitar a los estudiantes tanto la comprensión de las principales ideas de la estadística como la adquisición de una serie de habilidades útiles para trabajar con datos. Los ejemplos y los ejercicios que se presentan, aunque están pensados para principiantes, utilizan datos reales y dan una información general que permite a los estudiantes comprender el significado de sus cálculos. A menudo pido a los alumnos conclusiones que son algo más que un número (o “rechazar H_0 ”). Aparte de hacer cálculos y sacar conclusiones correctas o erróneas, algunos ejercicios exigen hacer valoraciones. Espero que los profesores fomenten en clase un amplio debate sobre los resultados de los ejercicios.

1. Ayudar a pensar como estadísticos:
 - (a) La necesidad de tener datos.
 - (b) La importancia de la obtención de datos.
 - (c) La omnipresencia de la variabilidad.
 - (d) La medición y calibración de la variabilidad.
2. Ofrecer más datos y conceptos, y menos teoría y fórmulas. Siempre que sea posible automatiza los cálculos y las representaciones gráficas. Un curso introductorio debe:
 - (a) Confiar plenamente en datos reales (no que solamente lo parezcan).
 - (b) Enfatizar los conceptos estadísticos; por ejemplo, causalidad frente a asociación, experimental *versus* observacional.
 - (c) Apoyarse más en los ordenadores que en las fórmulas de cálculo.
 - (d) No dar demasiada importancia a las demostraciones matemáticas.
3. Fomentar la enseñanza activa, a través de las siguientes alternativas al estudio individual:
 - (a) Discutir y resolver problemas en grupo.
 - (b) Hacer ejercicios con ordenadores.
 - (c) Trabajar con datos obtenidos en clase.
 - (d) Presentar ejercicios por escrito y exponerlos en el aula.
 - (e) Diseñar proyectos estadísticos de forma individual y en grupo.

Figura 1. Recomendaciones de la Sociedad Americana de Estadística (ASA) y de la Asociación Americana de Matemáticas (MAA).

Los capítulos 1 y 2 presentan los métodos y las ideas que unifican el análisis de datos. Los estudiantes agradecen la utilidad del análisis de datos, y el hecho de que realmente puedan hacerlo alivia un poco su inquietud sobre la estadística. Espero que los estudiantes se acostumbren a examinar los datos y que sigan realizando este análisis aun cuando el objetivo principal del análisis sea dar respuesta a una pregunta concreta mediante inferencia. El capítulo 3 trata sobre el muestreo aleatorio y los experimentos comparativos aleatorizados. Éstos, pese a encontrarse entre las ideas más importantes de la estadística, suelen omitirse injustamente al iniciar su enseñanza. El capítulo 4 se basa en las ideas del capítulo 3 y en las herramientas del análisis de datos del capítulo 1 para presentar el concepto de distribución muestral y, de una manera informal, el lenguaje de la probabilidad. El capítulo 5, que es opcional, presenta material adicional para cursos que exijan presentar la probabilidad de una manera más formal. El capítulo 6, que describe los razonamientos en los que se basa la inferencia estadística, es la piedra angular del resto del libro. Los restantes capítulos presentan métodos de inferencia aplicables a diversas situaciones, dando especial énfasis a los aspectos prácticos de la utilización de dichos métodos. Los capítulos 7 y 8 presentan los procedimientos básicos para hacer inferencia a partir de una y de dos muestras. Los capítulos 9, 10 y 11 (que se pueden leer de forma independiente y en cualquier orden) ofrecen una selección de algunos temas más avanzados. El capítulo 12 es una introducción a los métodos inferenciales no paramétricos.

Tecnología

Los cálculos automáticos aumentan la capacidad de los estudiantes para resolver problemas, reducen su frustración y les ayudan a concentrarse en las ideas y en la identificación del problema más que en la mecánica de su resolución. Este libro exige de los estudiantes que dispongan de una calculadora que pueda hallar la correlación y la regresión lineal simple, así como la media y la desviación típica. Como los estudiantes tienen calculadoras, el texto no discute las fórmulas de cálculo de la desviación típica muestral o de la recta de regresión mínimo-cuadrática. Los programas estadísticos tienen considerables ventajas sobre las calculadoras: la introducción y la edición de los datos es más fácil, y los gráficos son mucho mejores. Animo al empleo de programas estadísticos siempre que los recursos y el tiempo disponible para desarrollar el curso lo permitan. Este libro, sin embargo, no da por supuesto que los estudiantes utilizarán estos programas.

En el texto aparecen resultados obtenidos utilizando diversas tecnologías como por ejemplo calculadoras avanzadas, hojas de cálculo o programas estadísticos como el Minitab. Utilizar diversas tecnologías se ha hecho deliberadamente: los estudiantes tienen que entender que los procedimientos estadísticos que aparecen en el libro son universales. Teniendo conocimientos básicos de estadística se pueden leer e interpretar muchos resultados independientemente de la tecnología utilizada.

Texto accesible

Estadística aplicada básica intenta ser un texto moderno y accesible. En comparación con *Introduction to the Practice of Statistics* (IPS)³, *Estadística aplicada básica* guía más al lector, es más concreto y presenta menos material opcional. Aunque *Estadística aplicada básica* es de más fácil comprensión que IPS, sin embargo, los contenidos básicos son los mismos en ambos libros. Asimismo, en ambos libros se da mucha importancia tanto a trabajar con datos como a los problemas típicamente estadísticos. Los ejercicios que aparecen en las secciones “Aplica tus conocimientos” que aparecen situadas, después de la introducción de cada nuevo concepto, permiten valorar al propio alumno si ha asimilado el capítulo. Cada capítulo acaba con la sección “Repaso del capítulo”, un listado de sus contenidos básicos acompañado a menudo de un resumen gráfico de los principales conceptos.

¿Por qué lo hiciste de esta manera?

No existe una sola manera de organizar los temas en un curso de estadística para principiantes. En este libro su organización no se ha hecho de forma caprichosa. He aquí la respuesta a una serie de preguntas sobre el orden del contenido de este libro que se me han formulado con cierta frecuencia.

¿Por qué razón la distinción entre población y muestra no aparece en los capítulos 1 y 2? La distinción entre población y muestra no aparece en los dos

³David S. Moore y George P. McCabe, *Introduction to the Practice of Statistics*, 3ª ed., W. H. Freeman, New York, 1999.

primeros capítulos para dejar entrever que la estadística no es solamente inferencia. De hecho, la utilización de la inferencia estadística sólo es adecuada en determinadas circunstancias. Los capítulos 1 y 2 presentan las tácticas y herramientas utilizadas para describir cualquier tipo de datos. Estas tácticas y herramientas no dependen de la idea de inferencia que surge al pasar de muestra a población. Muchos conjuntos de datos de estos capítulos (por ejemplo los datos sobre los Estados europeos o los Estados norteamericanos) no deben nada a la inferencia ya que ellos mismos constituyen toda la población. John Tukey de los Laboratorios Bell y de la Universidad de Princeton, el filósofo del análisis de datos moderno, insiste en que cuando no sea relevante para el análisis, hay que evitar distinguir entre población y muestra. Tukey utiliza la palabra “grupo” (en inglés *batch*) para referirse, en general, a conjuntos de datos. Aunque no veo la necesidad de utilizar una palabra especial, sin embargo creo que Tukey tiene razón.

¿Por qué no empezar por la obtención de datos? De hecho, sería razonable empezar por la obtención de datos, ya que es el camino natural: cuando se planifica un estudio estadístico, se empieza por el diseño de la obtención de datos y se acaba con la inferencia. He escogido ubicar el diseño de la obtención de datos (capítulo 3) después del análisis de datos (capítulos 1 y 2) para enfatizar que las técnicas de análisis de datos se pueden aplicar a cualquier conjunto de datos. Uno de los principales objetivos del diseño estadístico de obtención de datos es hacer posible la inferencia. Por tanto, los contenidos del capítulo 3 son el camino natural hacia la inferencia.

¿Por qué las distribuciones normales aparecen en el capítulo 1? Las curvas de densidad en general, y entre ellas las normales, son una herramienta más para la descripción de variables cuantitativas, junto con los diagramas de tallos, los histogramas y los diagramas de caja. Cada vez es más frecuente que los programas estadísticos ofrezcan la posibilidad de calcular curvas de densidad además de, por ejemplo, dibujar histogramas. Personalmente, prefiero salirme de la presentación tradicional de las curvas de densidad y no presentarlas como algo absolutamente ligado a la probabilidad. Quiero que los estudiantes sigan el camino natural que parte de las representaciones gráficas, pasa por los resúmenes numéricos y llega a modelos matemáticos. Las curvas de densidad son un modelo habitualmente utilizado para describir de forma general la distribución de una variable. Finalmente, me gustaría contribuir a facilitar la *digestibilidad* de la probabilidad, un tema que tanto preocupa a los estudiantes. Trabajar pronto con distribuciones normales facilita la digestión posterior de los conceptos de

probabilidad, ya que ayuda a los estudiantes a ver las distribuciones de probabilidad como si fueran distribuciones de datos.

¿Por qué no presentar la correlación y la regresión más tarde, tal como es tradicional? Este libro empieza dando a los estudiantes la oportunidad de trabajar con datos. Al inicio del libro se estudia la estructura conceptual de esta parte de la estadística poco matemática pero muy importante. Además de trabajar con una sola variable, se ofrece a los estudiantes la posibilidad de analizar la relación entre varias variables. Es más, la correlación y la regresión vistas como herramientas descriptivas (capítulo 2) tienen un campo de actuación más amplio que la sola consideración de la inferencia (capítulo 11). Por ejemplo, la muy importante discusión sobre variables latentes tiene poco que ver con la inferencia. Personalmente, considero que el capítulo 2 es esencial, mientras que el capítulo 11 es opcional.

¿Dónde quedó toda la probabilidad? La mayor parte de la probabilidad quedó en el capítulo 5, que es optativo. A los estudiantes no les importa saltarse un capítulo, en cambio, a veces, les molestan los temas optativos dentro de los capítulos. En consecuencia, en esta nueva edición todo el capítulo 4 se considera obligatorio. En cambio, el capítulo 5 se deja como optativo.

Los profesores experimentados sabemos que los estudiantes encuentran difícil la probabilidad. La investigación sobre el aprendizaje confirma nuestra experiencia. Incluso entre los estudiantes que pueden resolver problemas formales de probabilidad, a menudo muestran una muy frágil interiorización de los conceptos probabilísticos. Intentar presentar una introducción sustancial a la probabilidad en un curso de estadística orientado a datos con un nivel asequible para estudiantes que no tengan una especial preparación en matemáticas, creo que no tiene mucho sentido. La probabilidad no ayuda a estos estudiantes a conocer en profundidad las ideas sobre inferencia (al menos, no tanto como los profesores imaginamos) y consume energías mentales que podrían aplicarse mejor a ideas esencialmente estadísticas.

En consecuencia, he presentado muy poca teoría de probabilidad en el capítulo 4. El capítulo conduce de forma directa a la idea de probabilidad como regularidad que aparece después de muchas repeticiones. A continuación, vemos formas concretas de asignar probabilidades y, finalmente, llegamos a la idea de distribución muestral de un estadístico. La ley de los grandes números y el teorema del límite central aparecen cuando se trata sobre la distribución de la media muestral. En este capítulo se omiten las *reglas generales de probabilidad*, así como la

probabilidad condicional y la combinatoria. Las reglas generales aparecen en el capítulo 5. La combinatoria es otro tema (incluso más difícil).

¿Por qué no has introducido un tema determinado? Los textos de introducción no tienen que ser enciclopedias. Introducir los posibles temas favoritos de distintos profesores tiene como consecuencia aumentar mucho el volumen del libro lo que puede asustar a los estudiantes. *Estadística aplicada básica* contiene básicamente temas de análisis de datos y de inferencia estadística, los campos más utilizados en la práctica, adecuados para comprender los conceptos estadísticos más importantes. Estos dos campos de la estadística se utilizan en muchos campos de aplicación. Emerson y Colditz⁴ comentan que con la estadística descriptiva, los procedimientos t y las tablas de contingencia, se puede comprender perfectamente la estadística del 73% de los artículos publicados en el *New Journal of Medicine*.

Agradecimientos

Quiero mostrar mi agradecimiento a mis colegas de distintas universidades por los comentarios que realizaron sobre los sucesivos borradores del manuscrito de este libro:

Elizabeth Applebaum
Park College

Steve Marsden
Glendale College

Edgar Avelino
Langara College

Darcy Mays
Virginia Commonwealth University

Smiley Cheng
University of Manitoba

Amy Salvati
Adirondack Community College

James Curl
Modesto Junior College

N. Paul Schembari
East Stroudsburg University

⁴J. D. Emerson y G. A. Colditz, "Use of statistical analysis in the *New England Journal of Medicine*", en *Medical Uses of Statistics*, 2^a ed., J. C. Bailar y F. Mosteller (eds.), NEJM Press, Boston, 1992, págs. 45-57.

David Gurney
Southeastern Louisiana University

W. Robert Stephenson
Iowa State University

Donald Harden
Georgia State University

Martin Tanner
Northwestern University

Sue Holt
Cabrillo College

Bruce Torrence
Randolph-Macon College

Elizabeth Houseworth
University of Oregon

Mike Turegon
Oklahoma City Community College

T. Henry Jablonski, Jr.
Eastern Tennessee State University

Jean Werner
Mansfield University of Pennsylvania

Tom Kaupe
Shoreline Community College

Dex Whittinghill
Rowan University

James Lang
Valencia Community College

Rodney Wong
University of California at Davis

Donald Loftgaarden
University of Montana

También quiero dar las gracias a quienes revisaron la primera edición:

Douglas M. Andrews
Wittenberg University

Ronald La Porte
Macomb Community College

Rebecca Busam
The Ohio State University

Ken McDonald
Northwest Missouri State University

Michael Butler
College of the Redwoods

William Notz
The Ohio State University

Carolyn Pilers Dobler
Gustavus Adolphus College

Mary Parker
Austin Community College

Joel B. Greenhouse
Carnegie Mellon University

Kenneth Ross
University of Oregon

Larry Griffey
Florida Community College at Jacksonville

Calvin Schmall
Solano Community College

Brenda Gunderson
The University of Michigan

Frank Soler
De Anza College

Catherine Cummins Hayes
University of Mobile

Linda Sorenson
Algoma University

Tim Hesterberg
Franklin & Marshall College

Tom Sutton
Mohawk College

Quiero agradecer de forma especial los comentarios de Sue Holt, surgidos después de utilizar durante mucho tiempo la primera edición de *Estadística aplicada básica* con estudiantes universitarios. También quiero dar gracias de forma especial a Patrick Farace, Diane Cimino Maass y, en general, a la gente de la editorial que ha contribuido de manera tan importante al diseño de la segunda edición en inglés.

Finalmente, estoy en deuda con muchos profesores de estadística con los que he discutido sobre la enseñanza de esta disciplina durante muchos años; con la gente de diversos campos de estudio con la que he trabajado en la comprensión de los datos; y, especialmente, con los estudiantes cuyas felicitaciones y quejas han cambiado y mejorado mi forma de enseñar. El trabajo con profesores de estadística, colegas de otras disciplinas y estudiantes me ha hecho recordar la importancia de la experiencia del trabajo directo con datos y de los razonamientos estadísticos en una época en la que los cálculos con ordenadores pueden hacer pasar por alto detalles estadísticamente importantes.

David S. Moore