

ENTENDIENDO LA ESTADÍSTICA: MODELOS, CONTROVERSIAS E INTERPRETACIONES

“En Dios confiamos, todos los demás traigan datos.”

W. E. Deming

**POR JOSÉ A. CRISTÓBAL
Y PILAR OLAVE**



Entendiendo la Estadística: modelos, controversias e interpretaciones

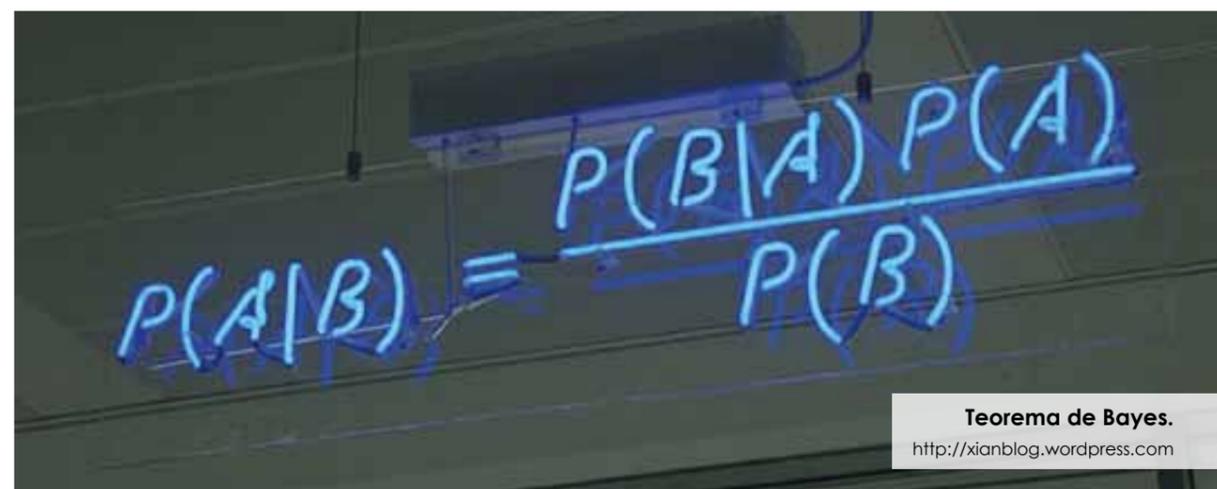
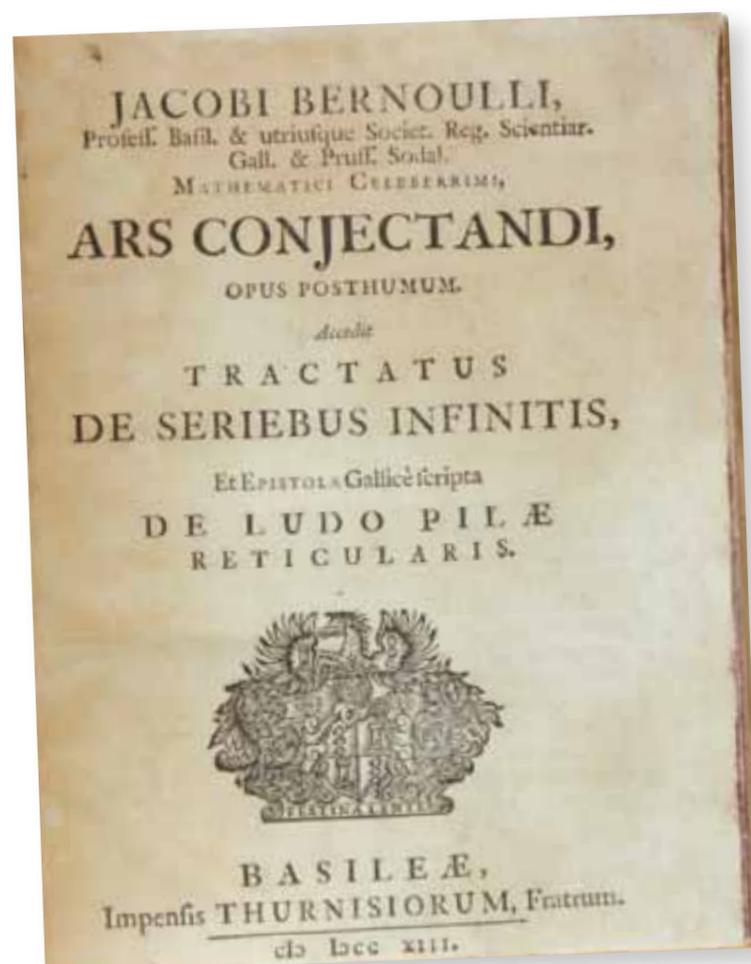
Este año 2013 celebramos el Año Internacional de la Estadística, coincidiendo con algunos importantes aniversarios, como la conmemoración de los 300 años de la publicación de "Ars Conjectandi" tratado pionero en el Cálculo de Probabilidades, y los 250 años de la publicación del Teorema de Bayes, un sencillo resultado analítico cuyas consecuencias en la creación de una corriente filosófica en la Estadística, con tanto auge como ha alcanzado en la actualidad, no podía ni imaginar su autor.

Con este hecho se pretende la aceptación, por parte del gran público, de un reconocimiento a la contribución que ha realizado la Estadística en el avance social y científico a lo largo de la Historia. Para ello, se han coordinado un elevado número de organizaciones de más de 100 países (entre ellas la Sociedad Española de Estadística e Investigación Operativa, distintas Universidades como la de Zaragoza, Oficinas de Administración Pública tales como el Instituto Nacional de Estadística o Instituto Aragonés de Estadística, medios de comunicación, etc.) con el fin de aumentar la conciencia ciudadana sobre la potencia e implicaciones de la Estadística en distintos aspectos de la sociedad, promocionar entre la juventud la oportunidad de dedicarse a la Estadística como una profesión, así como promover un incremento en la Investigación, Desarrollo e Innovación relacionadas con la Estadística.

.....
"Ars Conjectandi", 1713.

<http://www.vialibri.net>

En particular, se ha creado una página web: "www.statistics2013.org", en la que se detallan las actividades semanales realizadas en todo el mundo para conseguir los anteriores objetivos, y se incluyen enlaces a los distintos foros de discusión de cualquier tema relacionado con ellos. Por su parte, el Departamento de Métodos Estadísticos de nuestra Universidad de Zaragoza ha diseñado otra página web: "estadistica2013.unizar.es" con la misma finalidad, poniendo un especial énfasis en los acontecimientos que tienen lugar en España, y ha puesto en marcha una serie interactiva de "desafíos estadísticos" que aparecen en el suplemento Tercer Milenio del diario "Heraldo de Aragón", así como en el blog de ciencia de dicho periódico. Además, se ha difundido un vídeo promo-



cional de la Estadística y su importancia en la consecución del estado de bienestar de la humanidad (dada la relevancia de la utilización de datos estadísticos en la toma de decisiones de los Agentes Sociales), el cual puede reproducirse en YouTube y, si se desea, con subtítulos en español.

En este contexto, hemos considerado esta revista de la Facultad de Ciencias un foro adecuado para difundir las actividades anteriormente citadas, así como desarrollar unas breves ideas sobre cómo se puede adaptar un modelo a una situación aleatoria y la interpretación de los resultados que se deducen de él. Además, incluimos algunas de las más célebres controversias sobre esas interpretaciones, que pueden servir como aliciente al lector para seguir buscando otras más complejas, afianzando a la vez un modo de pensar llamado *razonamiento estadístico*.

MODELOS ESTADÍSTICOS

Todo el mundo acepta la necesidad de plantear modelos abstractos para simplificar situaciones físicas complicadas, y permitir extraer algunas conclusiones que de otro modo sería mucho más complejo o incluso inviable. Así,

cuando deseamos saber cómo se comporta el movimiento de los astros en el espacio, asociamos a esta situación un modelo matemático basado en una cierta definición de fuerzas y un conjunto de simplificaciones físicas que, mediante la resolución de unos sistemas de ecuaciones nos permite predecir, por ejemplo, dónde estará un determinado astro en un instante fijo.

Cuando se trata de situaciones que llevan asociada una cierta carga de incertidumbre, esto es, donde intervienen variables cuyos valores dependen de la ocurrencia o no de algún fenómeno de naturaleza aleatoria, también podemos incorporar otro tipo de modelos -denominados *estocásticos*- que permiten analizar cómo se comporta la incertidumbre de esos fenómenos aleatorios.

Imaginemos que deseamos establecer la relación entre el consumo privado y la renta disponible por familia en España en un determinado periodo de tiempo. Podemos crear un modelo que ligue ambas variables aleatorias (habría que incorporar también las demás variables que pueden estar relacionadas con ambas, así como la especificación de cómo se reparte la carga de incertidumbre sobre ellas, lo que se conoce como su distribución

Entendiendo la Estadística: modelos, controversias e interpretaciones

de probabilidad conjunta) para predecir, por ejemplo, el valor medio del consumo privado familiar cuando la renta queda dentro de un intervalo determinado. O bien, en diseños de Política Macroeconómica, cuando queremos saber el impacto sobre la inflación en la zona Euro de un incremento del tipo de interés nominal (en el que el Banco Central Europeo sí puede intervenir), se hace necesario plantear un modelo y estimar los parámetros asociados, cuyos valores permiten interpretar la relevancia del control del BCE en las expectativas de inflación.

Para que tales modelos sean útiles, se deben contrastar los resultados que se derivan de ellos con sus análogos en la realidad observable, y de esto se ocupa la rama de la Estadística conocida con el nombre de *Inferencia Estadística*. La manera de realizarlo (denominado *contraste* o *test de hipótesis*) consiste en comparar dichos resultados teóricos con las correspondientes medidas empíricas (observadas) en una cierta muestra obtenida aleatoriamente a partir de la población que se estudia. Por lo tanto, el lenguaje en el que se basa la Estadística tiene que venir dado forzosamente en términos de probabilidades.

Naturalmente, aunque el modelo sea completamente fiel, se observarán diferencias distintas de cero entre ambos valores, teóricos y empíricos, debido a su naturaleza aleatoria. La Estadística se ocupa de utilizar los valores de la muestra para aceptar o rechazar el modelo propuesto o, en su caso, realizar estimaciones sobre algún parámetro u otra característica que falte por conocer en el modelo. De este modo, mientras en el Cálculo de Probabilidades, para un modelo estocástico fijado, el objetivo es ¿qué se puede decir sobre los resultados?, en la Inferencia Estadística se parte de un conjunto de resultados y el problema es ¿qué se puede decir sobre el modelo probabilístico?

Hay situaciones en las que los modelos se pueden ir modificando para mejorar su ajuste, pero como el objetivo de construir un modelo es el de entender con mayor facilidad el fenómeno (aleatorio) que genera los datos observados, no necesita estar completo ni establecer de modo exacto todas las relaciones. Es famosa la cita del estadístico y matemático G. E. P. Box, fallecido este mismo año: *"Todos los modelos son incorrectos. Pero algunos son útiles"*.

Se pueden modelizar (recientemente, en un avance de la 23ª edición del Diccionario de la Real Academia Española se permite utilizar este término -en lugar de modelar- en la acepción de construir un esquema teórico) situaciones tan variadas y numerosas como el flujo de tráfico en una red de autopistas, la relación entre índices financieros y ciertas variables económicas, el control de una red de conexiones telefónicas, la mutación de genes a lo largo de varias generaciones, las ventas de los artículos producidos por una empresa en diversos segmentos de población, la duración del tiempo de desempleo en un país y su relación con ciertas variables socio-económicas, el control de calidad de un cierto proceso de producción...

SIMULACIÓN DE MODELOS: UNA HERRAMIENTA IMPRESCINDIBLE

Desde que poseemos una gran capacidad computacional debido al explosivo desarrollo de la tecnología informática, disponemos de una técnica para simular la ocurrencia de fenómenos aleatorios mediante un procedimiento basado en la extracción de "muestras artificiales", generadas con el ordenador a partir de las distribuciones de probabilidad del modelo. Estas muestras se llaman también *pseudo-aleatorias* porque en realidad están construidas mediante algoritmos deterministas (en principio basados en generadores congruenciales, aunque posteriormente se vienen utilizando otros mucho

más sofisticados), pero sus valores tienen frecuencias compatibles con las distribuciones de probabilidad del modelo. Para un lector no familiarizado con estas técnicas puede resultar paradójico que se utilicen herramientas no aleatorias para obtener unos datos que puedan considerarse resultado de un proceso aleatorio. Sin embargo, hay que tener en cuenta que es necesario realizar previamente un cuidadoso análisis matemático para poder aceptar, con una elevada confianza, que un algoritmo dado genere números que sean suficientemente "aleatorios" para ser útiles en el análisis del fenómeno aleatorio de interés. Es conocida la frase de R. R. Coveyou: *"La generación de números aleatorios es demasiado importante como para dejarlo al azar"*.

Ahora bien, hay que destacar una diferencia fundamental entre el modelo para una situación real y el modelo simulado. Habitualmente, de una situación aleatoria podemos extraer muestras una sola vez. Así, cuando observamos una muestra de 500 individuos de una cierta ciudad para observar en ellos la renta obtenida en el

"Todo el mundo acepta la necesidad de plantear modelos abstractos para simplificar situaciones físicas complicadas y permitir extraer algunas conclusiones que, de otro modo, sería mucho más complejo o incluso inviable."

Entendiendo la Estadística: modelos, controversias e interpretaciones

año anterior, es usual que no podamos repetir la experimentación. Pero cuando simulamos el modelo, podemos extraer tantas muestras de tamaño 500 como deseemos repitiendo el proceso de simulación. Esto hace que si, por ejemplo, nos interesa el valor medio de la muestra, en el caso de la situación real solo tenemos un valor de este indicador y nada podemos decir de la variabilidad de dicho indicador (recordemos que es aleatorio porque se ha obtenido mediante un procedimiento aleatorio). Sin embargo, el modelo simulado lo podemos repetir por ejemplo 1.000 veces (el cómputo electrónico es realmente barato y rápido), y estos valores nos dan información sobre la variabilidad del indicador que, en definitiva, son los utilizados para calibrar su utilidad y precisión.

ALGUNAS CONTROVERSIAS EN ESTADÍSTICA

Errores en la intuición sobre las regularidades en grandes muestras

La Estadística se basa en realizar mediciones de una característica sobre una muestra para hacer inferencias de los valores de esa característica en toda la población. Pero la intuición sobre el efecto del tamaño de la muestra presenta muchas veces ciertas dificultades en las personas que no son expertas en Estadística. Un ejemplo de este hecho es el conocido problema propuesto por D. Kahneman y A. Tversky a un amplio grupo de personas:

En una ciudad hay dos hospitales de distinto tamaño. En el mayor de ellos nace una media de 50 bebés al día mientras que en el menor la media es de 15. Podemos aceptar que, aproximadamente, la mitad

de los bebés que nacen son niños y la otra mitad niñas. Se elige un determinado día al azar y se observa que el 60% de los bebés nacidos en uno de los hospitales son niños. ¿En qué hospital es más probable que haya ocurrido este hecho?

Los autores citados publicaron en 1972 un tratado con los resultados de su experimento: muchos sujetos contestaban que, puesto que las dos variables aleatorias que miden el sexo de los nacidos en cada hospital son iguales, es igual de verosímil que los datos provengan de cualquiera de los dos hospitales. Pero otras personas se dieron cuenta de que, al aproximar la probabilidad de un resultado tras una repetición finita de veces por lo que sucedería en el caso límite, la aproximación será mejor, en general, cuanto mayor sea el número de repeticiones. De este modo, cuanto mayor sea el tamaño del hospital, el porcentaje de niños varones debe acercarse más al valor de la probabilidad de que un neonato sea varón. Así que la respuesta a la anterior pregunta es que los datos provienen más probablemente del hospital pequeño puesto que, al haber un número menor de repeticiones, es de esperar una mayor variabilidad del porcentaje de niños respecto del valor ideal del 50%.

“La intuición sobre el efecto del tamaño de la muestra presenta muchas veces ciertas dificultades en las personas que no son expertas en Estadística.”

En general estos investigadores encontraron que los individuos tienden a sobrevalorar la velocidad de convergencia expuesta en las *leyes de los grandes números* (aproximación de las frecuencias a la probabilidad). De este modo, cuando se lanza una moneda 8 veces y sale cara en tan solo 2 de ellas, muchas personas tienden a pensar que es casi imposible que la moneda esté equilibrada, cuando



<http://www.maderoeventos.com>

la probabilidad de obtener 2 caras en 8 lanzamientos de una moneda legal es del 11% y, por lo tanto, no es tan pequeña.

Errores en la interpretación de una probabilidad condicionada

Dada la creciente aplicación de la Probabilidad en los diversos campos de la vida cotidiana, debe ponerse atención para no cometer errores de interpretación que se dan con cierta frecuencia al tratar probabilidades condicionadas. Los genetistas forenses resaltan entre otras la llamada “falacia del fiscal” (o falacia del defensor, según su objetivo).

Tomemos en particular un caso publicado en el diario “El País” en 2003 sobre el resultado del análisis del ADN en restos hallados en el cuer-

po de S. Carabantes, en el que se afirmaba que la probabilidad de que pertenecieran a la misma persona que dejó sus restos genéticos junto a otro cadáver era de 99.999997%, por lo que parece lógico aceptar que se trata de la misma persona. Pero al parecer, la probabilidad anterior había sido calculada como la complementaria de $P(A/G) = 0.000003$, o probabilidad del suceso de que, fijado un cierto ADN (suceso G), una persona de la población tomada al azar (suceso A) lo tenga. Sin embargo, se interpretaba como la probabilidad $p = P(G/A)$ de que, dada una persona cogida al azar en la población, tuviese ese ADN observado. Este valor podría ser muy diferente del anterior, según fuera la probabilidad no condicionada, la cual es necesaria conocer para calcular p utilizando el conocido teorema de Bayes.

Entendiendo la Estadística: modelos, controversias e interpretaciones

Este error al confundir la probabilidad de ser inocente condicionado a una evidencia con la de obtener una evidencia en un individuo inocente es lo que se conoce, en términos generales, como la falacia del fiscal, y hay casos en la literatura forense en los que ha sido motivo de anulación de una primera sentencia. Quizá el primer ejemplo conocido sea el de "El pueblo contra Collins", correspondiente a un caso de asalto a una mujer en Los Ángeles en 1964, en los que la evidencia estaba basada en características físicas del sospechoso y su vehículo. Otro caso muy nombrado es el de una mujer, S. Clark, que tuvo un hijo que murió antes de cumplir el año, informando que se debía a una muerte súbita; pero un segundo hijo también murió en edad temprana por la misma causa, y la madre fue acusada de asfixiar a ambos, utilizándose como prueba una probabilidad errónea del tipo arriba citado. Tras una revisión del caso tres años más tarde fue declarada inocente. Otros muchos casos con historias similares se detallan, por ejemplo, en la página web "www.historiasdelaciencia.com" y en los enlaces que allí aparecen.

Esta misma falacia, en la que se confunde una probabilidad condicionada con la probabilidad condicionada traspuesta se da no solo en el área jurídica. Por ejemplo, en Medicina se presenta este mismo fenómeno cuando se sustituye el suceso ser inocente con el de poseer una determinada enfermedad y el de encontrar una evidencia con la obtención de un positivo en un test de hipótesis para detectarla. En esta situación, la probabilidad condicionada se convierte en la probabilidad de falso positivo (en la actualidad ha aparecido frecuentemente en los medios de comunicación el caso del elevado número de falsos positivos que se presenta en la detección de tumores en las pruebas de mamografía).

EL MUESTREO, CLAVE EN LA INTERPRETACIÓN DE RESULTADOS

Cuando se realiza un muestreo aleatorio hay que distinguir el caso más sencillo en el cual los individuos se incorporan a la muestra todos ellos con la misma probabilidad, en cuyo caso se denomina *muestreo simple* o equiprobable, y aquél en donde las probabilidades de inclusión de los individuos en la muestra son diferentes, en cuyo caso se obtiene lo que se denomina un *muestreo sesgado*. Por ello, las distribuciones de probabilidad de un mismo indicador estadístico, construido a partir de una muestra sesgada y otra que no lo es, ya no son iguales. Y aunque ambos son igualmente útiles si se toma en cada caso la distribución que le corresponde, debemos tener cuidado para no actuar erróneamente

La controversia del caso Sally Clark en un periódico de la época.

<http://www.neoteo.com>

desarrollando los cálculos como si la muestra fuera simple cuando en realidad está sesgada, pues hay situaciones donde el sesgo está "escondido".

Veamos el siguiente caso descrito por G. P. Patil (1984). En 1966 se realizó una encuesta en los hoteles de Marruecos para estimar la duración media de la estancia de los turistas extranjeros en el país, que resultó ser de 17,8 días. Pero también se hizo otra encuesta con el mismo objetivo contactando con los turistas a su salida del país, y resultó ser de 9,0 días, prácticamente la mitad de la anterior estimación. Estas cifras tan dispares, en realidad no suponen ninguna contradicción, pues están midiendo variables diferentes. En el primer caso los turistas encuestados forman una muestra sesgada, ya que al ser los instantes de visita a los hoteles aleatorios, los turistas que más frecuentemente encontraban los encuestadores eran aquellos cuya estancia era más larga, y por ello las personas que tenían estancias más largas eran recogidos en la muestra con mayor probabilidad. Sin embargo, en el segundo caso la muestra era simple.

El tipo de sesgo tratado en el ejemplo anterior se denomina *sesgo por longitud*, y supone que la probabilidad de inclusión en la muestra es proporcional al valor de la variable que se observa. Ha sido estudiado en la literatura especializada en las áreas de tipo tecnológico, por su vinculación a problemas estadísticos asociados a ciertos procesos que analizan la ocurrencia de una serie de sucesos, y los tiempos en que tienen lugar. Cox analizó un conocido problema en referencia a la longitud de unas fibras textiles (ver por ejemplo J. L. Ojeda y otros, (2011) con respecto a una breve exposición de situaciones en diversas áreas). Pero también tienen una importancia destacada en otros problemas económico-sociales, asociados a tiempos de duración, como puede



<http://www.eldebate.com.ar>

ser el estudio del tiempo de desempleo de una población, dentro del ámbito de los mercados laborales.

En efecto, si en un determinado instante de tiempo elegido al azar contactamos con una Oficina de Empleo y extraemos una muestra de los individuos que en ese momento están desempleados, tal muestra está también sesgada por longitud, pues aquellas personas que tienen duraciones de desempleo más largas son incluidas en la muestra con mayor probabilidad. Además, si solo podemos medir el tiempo desde que el individuo comenzó a estar desempleado hasta el momento del muestreo, nues-



Entendiendo la Estadística: modelos, controversias e interpretaciones

tras observaciones no son completas (no miden toda la duración del intervalo de desempleo) y constituyen lo que se denomina unos datos "censurados", que requieren un tratamiento estadístico especial.

La problemática asociada a los datos sesgados por longitud está relacionada con la "paradoja del tiempo de espera" descrita por W. Feller (1971): un observador en una parada de autobús puede registrar el tiempo que transcurre entre la llegada de dos autobuses consecutivos y estimar la media de esa distribución. Otro observador accede a la parada en un instante aleatorio (uniforme) en el periodo entre dos llegadas consecutivas de autobús y mide el tiempo que tiene que esperar hasta que llegue el primero de ellos. Pues bien, el tiempo medio que estime este último no tiene por qué ser la mitad del estimado por el primer observador, como pudiera parecer en un principio. Lo que ocurre es que el segundo ob-

servador mide datos sesgados, porque cuanto mayor es un intervalo entre llegadas, hay más probabilidad de que el observador llegue en ese intervalo.

Todo ello nos da una idea, tal como ya hemos dicho, de las dificultades que para interpretar bien los datos pueden tener profesionales de muy distintas áreas si no son expertos en realizar análisis estadísticos, como pueden ser juristas, periodistas y otras personas encargadas de transmitir y difundir información.

Por otra parte, ya hemos hecho alusión anteriormente al hecho de que usualmente tan solo se posee una muestra para realizar el análisis estadístico, y lo importante que sería poder disponer de un cierto número de ellas para poder hacer averiguaciones, no solo de la variabilidad de los indicadores que analicemos, sino en general de cómo está distribuida toda su probabilidad.



<http://www.zoomnews.es>



<http://nadandoenunmardedatos.blogspot.com.es>

B. Efron (1979) introdujo un nuevo método con este fin (aunque las ideas genéricas ya se habían establecido con anterioridad) que estaba basado en obtener sucesivas "remuestras" a partir de la muestra observada, mediante un muestreo independiente con reemplazamiento de la misma. Por lo tanto, queda justificado que, de la misma forma que se ha obtenido la muestra original, podrían haberse observado cualquiera de las remuestras: existe la misma relación entre la distribución del modelo original y la distribución de la muestra, que entre la distribución de la muestra y cualquiera de las remuestras.

Podemos pensar en un símil con tres muñecas rusas metidas cada una dentro de la anterior.

Si solo podemos levantar la primera muñeca, tenemos toda la información sobre la primera y segunda muñecas, pero no sobre la tercera. Ahora bien, si nos fijamos en una característica cualquiera (como por ejemplo la distancia entre los ojos de las muñecas) y suponemos una relación igual para esta característica entre cada par consecutivo, de los datos de las dos primeras podremos inferir el valor de la característica en la tercera.

A este método de remuestreo lo bautizó Efron con el nombre de "bootstrap", haciendo mención así a su carácter de autosuficiencia (se emplea una misma herramienta para una función más general que aquella para la que se ha concebido, sin utilizar ningún apoyo exter-

Entendiendo la Estadística: modelos, controversias e interpretaciones

no). En realidad, *bootstrap* es el término en inglés que representa las correas del calzado, y el origen de tal nombre para este método de remuestreo proviene del relato de las *Aventuras del barón de Münchhausen*, quien logra sacarse a sí mismo (*autosuficiente*) de una laguna a la que había caído tirando solamente de las correas de sus botas (*bootstrap*). Este personaje literario (basado en el auténtico barón, que vivió en el siglo XVIII) fue creado por R. E. Raspe, y se caracteriza por las hazañas extravagantes de las que fue protagonista (el lector interesado en ellas puede ver enlaces, por ejemplo, en la página web de Wikipedia).

“Todo ello nos da una idea de las dificultades que para interpretar bien los datos pueden tener profesionales de muy distintas áreas si no son expertos en realizar análisis estadísticos.”

Las ideas teóricas subyacentes en esta nueva filosofía inferencial son sencillas, pero requieren repetir, repetir... procedimientos computacionales sobre datos similares, por lo que son llamados métodos estadísticos computacionalmente intensivos. El número de remuestras necesario para estimar ciertas características de la distribución del indicador estadístico puede ser enormemente elevado, pero el desarrollo reciente del cálculo computacional ha dado a la Estadística una nueva perspectiva y ha permitido desarrollar técnicas como el *bootstrap* para sustituir un tratamiento teórico, con la ventaja además de que supone, en un gran número de casos, una mejor aproximación a la solución. En P. Olave, (2000) puede verse una introducción a este método de remuestreo y aplicaciones a algunos problemas relevantes en Economía, como por ejemplo la evolución de la población en las comarcas de Aragón, o en Economía Financiera para el cálculo de intervalos de predicción de la prima de riesgo en mercados de renta variable.



Finalmente, solo nos queda destacar una última reflexión en este Año Internacional de la Estadística. La interpretación correcta de ciertas aseveraciones estadísticas sobre los datos no es, como se ha visto, fácil en muchos contextos, y requiere de una lectura reposada e incluso de la ayuda de expertos para evitar consecuencias erróneas.

José A. Cristóbal

Miembro del Senatus Científico
Dpto. de Métodos Estadísticos
Facultad de Ciencias
Universidad de Zaragoza

Pilar Olave

Dpto. de Estructura e Historia Económica y Economía Pública
Facultad de Economía y Empresa
Universidad de Zaragoza

BIBLIOGRAFÍA

Box G. E. P. (1979), "Robustness in the Strategy of Scientific Model Building" in *Robustness in Statistics: Proceedings of a Workshop* (ed. by RL Launer and GN Wilkinson).

Coveyou R. R. (1998). En "The Jungles of Randomness: A Mathematical Safari" (I. Peterson), Wiley, pp. 178.

Efron B. (1979), "Bootstrap Methods: Another Look at the Jackknife". *Ann. Statist.* Vol 7, Num 1, 1-26.

Feller W. (1971), "An introduction to probability theory and its applications", Vol 2. Wiley.

Kahneman D. y Tversky A. (1974), "Judgment under uncertainty: Heuristics and biases" *Science*, 185, 1124-1131.

Ojeda J. L., Cristóbal J. A. y Alcalá J. T. (2011), "Biased data and its Analysis", *BEIO*, 27 (2), 91-101.

Olave P. (2000), "La Estadística Actual. El método Bootstrap", Cuadernos Económicos Escuela y Despensa, 10. Facultad de Economía y Empresa. Zaragoza.

Patil G. P. (1984), "Studies in statistical ecology involving weighted distributions. *Statistics: Applications and New Directions*", *Indian Stat. Inst., Calcutta*, 478-503.

