



APLICACIÓN DEL MÉTODO BOOTSTRAP AL CONTRASTE DE HIPÓTESIS EN LA INVESTIGACIÓN EDUCATIVA

JAVIER GIL FLORES (*)

RESUMEN. En este artículo presentamos el método Bootstrap, que desde su formalización por Efron (1979) ha llegado a constituirse en el más popular de los procedimientos de remuestreo. Tras revisar de manera introductoria los fundamentos del mismo, analizamos sus posibilidades para la inferencia estadística, y concretamente para el contraste de hipótesis. La escasez de aplicaciones en el ámbito de la investigación educativa nos lleva a presentar dos ejemplos de utilización que sirven al propósito de ilustrar el procedimiento y valorar este enfoque frente a los métodos clásicos de la estadística inferencial.

ABSTRACT. In this article we introduce the Bootstrap method, which since its invention by Efron (1979) has become the most popular of the resampling procedures. After an initial revision of its foundations, we analyse the possibilities for statistical inference and more specifically for hypothesis contrast. The lack of applications in the field of educational research leads us to present two examples of usage, which serve to both illustrate the procedure and recognise the importance of this approach compared to the more traditional methods of inferential statistics.

Posiblemente la aplicación de los métodos Bootstrap suponga uno de los avances más relevantes de la moderna estadística (Efron, 1995), a pesar de que los investigadores educativos de nuestro contexto inmediato no los hayan incorporado aún al análisis de datos que realizan en el curso de sus investigaciones. El método Bootstrap se enmarca entre los procedi-

mientos de remuestreo, consistentes en generar un elevado número de muestras como base para estudiar el comportamiento de determinados estadísticos. A nivel práctico, la actual facilidad para realizar procedimientos iterativos de manera informatizada elimina los posibles obstáculos que la aplicación de este tipo de métodos pudiera representar. En el

(*) Universidad de Sevilla.

presente trabajo trataremos de describir someramente los fundamentos de estos métodos (para profundizar en los mismos, pueden consultarse las obras de Efron, 1979, 1982; Hall, 1992; Efron y Tibshirani, 1993; Mooney y Duval, 1993; Shao y Tu, 1995; Davison y Hinkley, 1997; Chernick, 1999) y ejemplificaremos su utilización en dos situaciones reales de investigación, en las que aplicamos el enfoque Bootstrap al contraste de hipótesis.

BOOTSTRAP E INFERENCIA ESTADÍSTICA

Buena parte de la estadística convencional se apoya en la suposición de que las variables estudiadas se distribuyen de acuerdo con el modelo de la curva normal. Bajo ese supuesto, o recurriendo a leyes que permiten salvar la no-normalidad poblacional cuando el tamaño de las muestras es suficientemente grande, es posible determinar la distribución muestral de diferentes estadísticos que utilizamos en la estimación de parámetros poblacionales y en el contraste de hipótesis acerca de tales parámetros.

Por ejemplo, la comprobación de hipótesis estadísticas acerca del parámetro media se apoya en la suposición de que la distribución muestral del estadístico media sigue un modelo normal cuando la variable estudiada se distribuye normalmente en la población. Además, de acuerdo con el teorema del límite central, la distribución muestral del estadístico media es normal $N(\mu, s/\sqrt{n})$ cuando el tamaño muestral sea suficientemente grande (por acuerdo, de al menos $n=30$), incluso cuando no se cumpla la normalidad de la población. La distribución muestral de otros estadísticos sigue también modelos conocidos tales como la t de Student, la distribución F o la distribución² y disponemos de fórmulas para el

cálculo de los parámetros media y error típico de tales distribuciones.

Basándonos en los métodos Bootstrap, la idea de fondo sigue siendo la de construir un modelo de distribución para determinados estadísticos a partir de la información proporcionada por la muestra, aunque el modo de proceder es distinto. Como hemos señalado, en los métodos estadísticos clásicos la base para hacer inferencias sobre la población se encuentra en suponer para los estadísticos una distribución muestral teórica, cuyos parámetros pueden ser estimados a partir de estadísticos observados en la muestra. En cambio, los procedimientos basados en el Bootstrap implican obviar los supuestos sobre la distribución teórica que siguen los estadísticos. En su lugar, la distribución del estadístico se determina simulando un número elevado de muestras aleatorias construidas directamente a partir de los datos observados. Es decir, utilizamos la muestra original para generar a partir de ella nuevas muestras que sirvan de base para estimar inductivamente la forma de la distribución muestral de los estadísticos, en lugar de partir de una distribución teórica asumida *a priori*.

Este enfoque tiene su antecedente inmediato en las técnicas de simulación Monte Carlo, consistentes en extraer un número elevado de muestras aleatorias de una población conocida para calcular a partir de ellas el valor del estadístico cuya distribución muestral pretende ser estimada (Money, 1997). Sin embargo, en la práctica no solemos conocer la población y lo que manejamos es una muestra extraída de ella. El investigador parte de un conjunto de datos observados, que constituyen una muestra extraída de la población que pretende estudiar. Cuando las técnicas Monte Carlo son aplicadas a la resolución de problemas estadísticos,

partiendo de datos observados en una muestra, reciben más apropiadamente la denominación de «técnicas de remuestreo».

Existen diversos procedimientos utilizados para generar muestras a partir de un conjunto de observaciones y construir una distribución muestral que pueda ser usada para la estimación de intervalos de confianza y el contraste de hipótesis (véase una clasificación al respecto en Rodgers, 1999). Entre ellos, posiblemente el más conocido y comúnmente utilizado es el método Bootstrap. La primera alusión a este método bajo tal denominación se debe a Efron (1979), si bien las ideas básicas del mismo se venían manejando desde al menos una década atrás (Simon, 1969). Efron concibió el método Bootstrap como una explicación o ampliación de las «técnicas jackknife», las cuales suelen proceder extrayendo muestras construidas al suprimir cada vez un solo elemento de la muestra original para valorar el efecto sobre determinados estadísticos (Quenouille, 1949; Tukey, 1958; Miller, 1974).

Referido al caso de la estimación de parámetros, las técnicas Bootstrap implican básicamente desarrollar un proceso en el que distinguimos diferentes pasos:

- A partir de la muestra original $\{X_1, X_2, \dots, X_n\}$, se extrae una nueva muestra $\{X_1^*, X_2^*, \dots, X_n^*\}$, por medio de muestreo con reposición. Es decir, tras la extracción de un primer elemento, éste se repone en la muestra original de tal forma que podría ser elegido de nuevo como segundo elemento de la muestra extraída. De este modo, cada observación individual tiene una probabilidad $1/n$ de ser elegida cada vez, como si el muestreo se realizara sin reposición en un universo infinita-

mente grande construido a partir de la información que provee la muestra.

- Para la muestra obtenida se calcula el valor de un determinado estadístico $\hat{\theta}$ que se utiliza como estimador del parámetro poblacional θ , en cuyo estudio estamos interesados.
- Repetimos los dos pasos anteriores, hasta obtener un elevado número de estimaciones $\hat{\theta}^*$. En este punto, el recurso a herramientas informáticas que desarrollen las tareas de selección de muestras y determinación de las estimaciones resultará ineludible.
- Se construye una distribución empírica del estadístico $\hat{\theta}$, que representa una buena aproximación a la verdadera distribución de probabilidad para ese estadístico. Es decir, se determina de este modo la distribución muestral de un estadístico sin haber hecho suposiciones sobre la distribución teórica a la que ésta se ajusta y sin manejar fórmulas analíticas para determinar los correspondientes parámetros de esa distribución.

De acuerdo con la idea central en que se basa el método Bootstrap, el procedimiento supone utilizar la muestra considerando que en sí misma contiene la información básica sobre la población. Por tanto, la adecuación de este método será tanto mayor cuanto más información aporte la muestra sobre la población. Una consecuencia directa es que a medida que aumenta el tamaño de la muestra mejor será la estimación que podemos hacer sobre la distribución muestral de un

estadístico. No obstante, incluso con muestras pequeñas, entre 10 y 20 casos, el método Bootstrap puede ofrecer resultados correctos (Bickel y Krieger, 1989), juzgándose inadecuados para muestras de tamaño inferior a 5 (Chernick, 1999). Con un tamaño suficientemente grande, el incremento en el número de muestras procurará una mejora en la estimación de la distribución muestral.

El Bootstrap y los procedimientos de remuestreo en general, han comenzado a centrar la atención de los estadísticos a partir de la década de los ochenta, cuando el desarrollo de la informática allanó los obstáculos prácticos unidos a la simulación de un número elevado de muestras. A finales de esta década, la utilización del método Bootstrap para el contraste de hipótesis empezaba a ser considerada una alternativa a los tests paramétricos y no paramétricos convencionales (Noreen, 1989).

Como se advierte a partir de la breve presentación que hemos realizado, el método Bootstrap no representa un estadístico en sí mismo, sino un enfoque sobre cómo usar los estadísticos de cara a hacer inferencias acerca de los parámetros (Mooney y Duval, 1993). Más que una utilización de manera aislada, estos métodos se prestan a la combinación con otras técnicas estadísticas. Aplicaciones de los procedimientos Bootstrap a problemas de estimación, ajuste a una curva, contraste de hipótesis, regresión, clasificación o análisis de componentes principales, entre otras, son presentadas en el manual de Efron y Tibshirani (1993); y, por citar otros ejemplos, se han tratado aplicaciones a los coeficientes de correlación (Hans y Kolen, 1988; Knapp, Noblitt y Viragoontavan, 2000), al análisis factorial (Ichikawa y Konishi, 1995), las series temporales (Romo, 1994) o a los modelos de ecuaciones estructurales (Hancock y

Nevitt, 1999; Nevitt y Hancock, 2001; Raykov, 2001).

Los métodos Bootstrap han ganado popularidad y han llegado a ser considerados una herramienta general para el trabajo estadístico (Young, 1994). La atención de la comunidad estadística hacia el Bootstrap queda reflejada en el importante número de trabajos dedicados a justificar sus bases teóricas que han sido generados desde la aparición del trabajo de Efron en 1979. Según el mismo autor (Efron, 2000), hasta finales de siglo el número de artículos consagrados a este tema superaba el millar.

UTILIZACIÓN DEL BOOTSTRAP EN LA INVESTIGACIÓN EDUCATIVA

Desde nuestra posición como investigadores educativos, el interés sobre el Bootstrap se enmarca en una perspectiva aplicada. Una clara utilidad de estas técnicas está en la posibilidad de hacer frente a los problemas de inferencia estadística en las situaciones en las que los supuestos sobre la población son insostenibles y los métodos paramétricos tradicionales no resultan adecuados.

Aunque buena parte de las técnicas estadísticas inferenciales descansan sobre el supuesto de la normalidad poblacional, en la práctica pocas veces van precedidas de la aplicación de algún test para comprobar si los datos proceden de una población en la que las variables estudiadas se distribuyen normalmente. Lo habitual es dar por supuesto que tal condición se cumple. Sin embargo, en el ámbito de la investigación educativa, y de la investigación social en general, el supuesto de normalidad puede llegar a ser poco realista en algunos casos. Algunos estadísticos han denunciado el abuso que se hace de las leyes de probabilidad, y en especial de

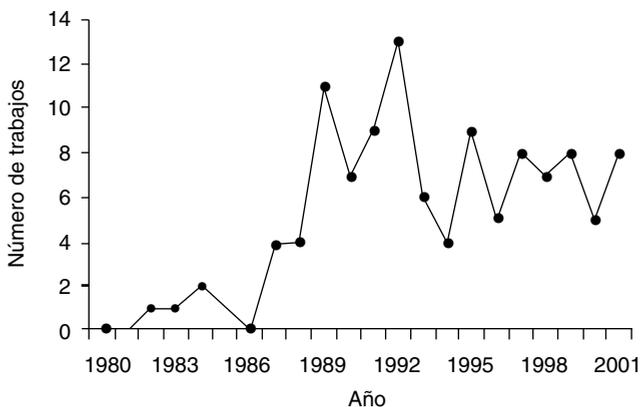
la ley normal. Sirvan como ejemplo las palabras de De Lagarde (1983, p. 2), para quien «en un universo aleatorio no está probado que puedan conocerse las leyes a las que obedecen los fenómenos, e incluso no es totalmente seguro que estas leyes existan realmente». Si ponemos en cuestión el supuesto de normalidad poblacional, no parece una mala estrategia determinar las características de la población a partir de una muestra que procede de ella. En realidad, a falta de un ajuste a modelos teóricos conocidos, ésta sería la mejor y única información posible a tener en cuenta.

Otra situación que aconsejaría el recurso a métodos como el Bootstrap se da cuando trabajamos con estadísticos para los cuales no se dispone de supuestos acerca de la forma adoptada por su distribución muestral. Es el caso, por ejemplo, de la diferencia de medianas o la asimetría, cuya distribución muestral es desconocida y, en consecuencia, no contamos con métodos paramétricos que permitan la inferencia estadística.

Al enumerar ámbitos en los que el Bootstrap ha sido aplicado, Chernick (1999, p. 7) incluía Psicología, Física,

Geología, Ecología, Ornitología, Econometría, Geología, Meteorología, Genética, Procesamiento de señales e imágenes, Medicina, Ingeniería, Química y Contabilidad, no estando incluida entre ellas la Educación. Para determinar la atención prestada a estos métodos y el uso hecho de los mismos en la investigación educativa, hemos recurrido a rastreos en las bases de datos que actúan como principales referentes en este campo. Concretamente, hemos explorado en la que posiblemente sea la principal base de datos sobre educación, elaborada por el *Educational Resources Information Center* (ERIC) y que recoge artículos de más de 750 revistas de este ámbito, así como informes y trabajos presentados a reuniones científicas, fundamentalmente en el ámbito norteamericano. Al recuperar los registros en los que se menciona a las técnicas Bootstrap, observamos en las dos últimas décadas un bajo número de trabajos, junto con una ligera tendencia ascendente (ver figura 1) que debe ser atenuada teniendo en cuenta el incremento que se ha registrado durante el mismo período en el número de publicaciones periódicas sobre educación.

FIGURA I
Presencia de trabajos sobre Bootstrap en la base de datos ERIC



En nuestro contexto más próximo, entre las bases de datos del Consejo Superior de Investigaciones Científicas se incluye ISOC, que recoge artículos publicados desde 1975 en revistas españolas de las áreas de Ciencias Sociales y Humanidades. Los trabajos que aluden al Bootstrap incluidos en esta base son un total de 11, fechados entre los años 1992 y 2001. La mayor parte de los mismos se encuadran en el ámbito de la Economía; sólo dos han sido publicados en revistas de Psicología, y ninguno proviene de revistas especializadas en Educación.

De esta breve exploración, cabe concluir que el uso que se hace de las técnicas Bootstrap en el ámbito de los estudios sobre educación, y particularmente en nuestro país, es escaso. Además, la mayoría de los trabajos se centran en el estudio de los métodos y sus fundamentos teóricos, siendo reducida la presencia de las técnicas Bootstrap en trabajos empíricos de investigación educativa. Tratando de aportar una pequeña contribución al conocimiento y difusión de los métodos Bootstrap, en las páginas que siguen mostraremos dos ejemplos de aplicación a situaciones reales de investigación educativa. Para ello, hemos elegido el contraste de hipótesis, y en particular el contraste de dos grupos, por tratarse de un problema estadístico que se plantean con frecuencia los investigadores educativos.

EJEMPLOS DE APLICACIÓN DEL BOOTSTRAP AL CONTRASTE DE DOS GRUPOS

Una situación familiar en los diseños de investigación, tanto experimentales como causal-comparativos, es la comparación de grupos a fin de determinar si existen diferencias significativas entre ellos. A esta situación corresponden los dos ejemplos que desarrollaremos, en los que se realizan sendos contrastes de hipótesis.

En el primero de ellos basaremos el contraste en el estadístico diferencia de medias, mientras que en el segundo caso recurriremos al cálculo del estadístico *t* de Student, que suele ser el habitualmente empleado desde enfoques paramétricos.

La implementación de este método requiere el recurso de herramientas informáticas. Los paquetes estadísticos de mayor difusión no suelen incluir procedimientos automatizados para realizar el Bootstrap, si bien algunos autores han apuntado modos en que podrían llevarse a cabo utilizando las posibilidades de programación que ofrecen paquetes como SPSS o SAS (Fan, 2001; Guthrie, 2001). Aquí utilizaremos el software *Resampling Stats Add-In for Excel* (Resampling Stats Inc., 2001), específicamente diseñado para aplicar procedimientos de remuestreo.

CONTRASTE BASADO EN LA DIFERENCIA DE MEDIAS

En el primero de los ejemplos de aplicación del Bootstrap al contraste de hipótesis sobre dos grupos nos basaremos en un sencillo ejemplo, en el que trabajamos con datos extraídos del estudio de Gil y Jaén (2001) sobre una estrategia docente utilizada en la enseñanza de los Métodos Cualitativos de Investigación Educativa a estudiantes de Pedagogía.

Tras utilizar una metodología didáctica basada en la realización de proyectos de investigación por parte de los alumnos, se quiso evaluar la experiencia recurriendo, entre otros aspectos, a los resultados logrados en términos de cambio de las actitudes hacia la investigación. Se utilizó una escala de medición de actitudes hacia la investigación cualitativa, administrada antes y después de la experiencia a fin de comprobar el modo en que podrían haberse modificado las actitudes de

partida en el grupo de alumnos que cursan la materia. La escala fue construida adaptando la utilizada por Chang (1996). En total, el instrumento constaba de 20 items o afirmaciones, ante los que el alumno debía manifestar su grado de acuerdo conforme a una escala de seis puntos.

Recogidos los datos, pudo calcularse la puntuación total en cada una de las aplicaciones del instrumento. La comparación entre la media alcanzada por el grupo de alumnos antes y después de la experiencia aportaría información sobre la posible modificación de actitudes hacia la investigación cualitativa, y en particular sobre la mejora de éstas.

Siguiendo el procedimiento convencional, habría que basarse en los valores promedio de ambos grupos, sometiendo a contraste la hipótesis nula de que son iguales los parámetros media en las poblaciones de las que fueron extraídas las respectivas muestras observadas. Bajo la hipótesis nula, el estadístico t , construido a partir de la diferencia de medias, se distribuiría siguiendo una distribución t de Student para un número determinado de grados de libertad. Si el valor observado se aleja suficientemente del valor medio de la distribución, que según la hipótesis nula es cero, podríamos rechazar el supuesto de partida.

La aplicación de un contraste basado en el estadístico t de Student requeriría la comprobación de supuestos previos tales como la normalidad de las variables estudiadas. Cuando la distribución del estadístico de contraste no sigue el modelo de la t , bien porque la variable estudiada no se distribuye normalmente en la población o bien porque el tamaño muestral resulta insuficiente, podríamos estar cometiendo un error en la inferencia.

En la situación que tomamos como ejemplo, la asignatura en la que se desarrolló la experiencia contaba con tan

sólo seis alumnos matriculados, dado su carácter optativo cuatrimestral, la existencia de horario único en turno de tarde y la presencia de otras materias optativas que atraen el interés del alumnado. En la primera aplicación de la escala se recogieron respuestas para los seis alumnos, mientras que en la aplicación final sólo pudo calcularse la puntuación total para cinco alumnos, dado que uno de ellos no respondió a la totalidad de los items. Consideraremos que se trata de muestras no relacionadas, dado que el carácter anónimo de la escala impediría establecer una correspondencia entre las escalas previa y posterior a la intervención. Así pues, la información de partida sería la proporcionada por la muestra $z = \{68, 73, 76, 80, 85, 92\}$ y la muestra $y = \{68, 88, 94, 100, 108\}$, cuyas medias son respectivamente 79 y 91,6.

Ante grupos tan pequeños, la suposición de normalidad parece arriesgada. En tal situación, una solución no paramétrica adecuada sería aproximarse al problema a través del método Bootstrap, construyendo nuestro propio modelo para valorar la diferencia de medias observada. La aplicación del Bootstrap supondría considerar que ambas muestras (z, y), proceden de poblaciones (F, G) posiblemente diferentes, y someter a contraste la hipótesis nula $H_0: F = G$. Para el contraste, utilizaremos el estadístico diferencia de medias, que en este caso alcanza un valor observado $\bar{y} - \bar{z} = 12,1$, y estudiaremos el modo en que se distribuye tal estadístico en un conjunto elevado de muestras generadas por remuestreo. El algoritmo que seguiremos es el siguiente:

- Si H_0 es cierta, z e y proceden de una población común. Para generar muestras de esa población común, utilizaremos la información disponible, que no es otra que la proporcionada por ambas

muestras. Así pues, combinando los elementos de z e y , obtenemos $x = \{z, y\}$ que constará de los 11 valores registrados. A partir de x , extraemos con reposición una muestra de tamaño $n=11$, denominando z^* a los 6 primeros valores, e y^* a los 5 restantes.

- Calculamos para la muestra extraída el valor del estadístico $\bar{y}^* - \bar{z}^*$, en el que basaremos el contraste de hipótesis.
- Repetimos los pasos anteriores un número elevado de veces. El número de repeticiones, y por tanto de muestras extraídas, será de 5.000

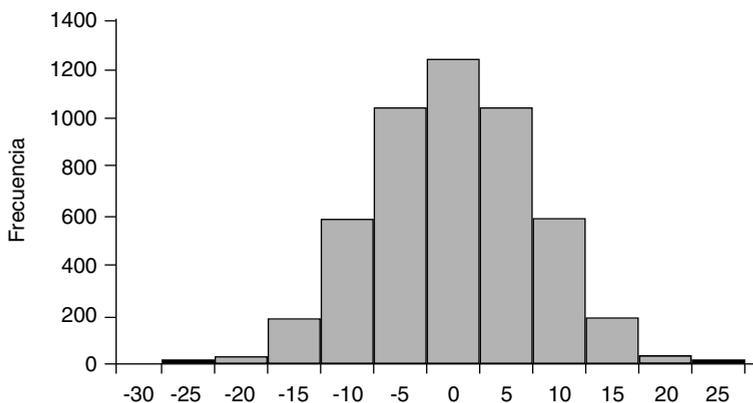
siguiendo las recomendaciones de Chernick (1999, p. 114), que considera esta cifra adecuada para la mayor parte de los problemas planteados.

- Construimos la distribución de frecuencias para el estadístico $\bar{y}^* - \bar{z}^*$. En la tabla I se recogen frecuencias absolutas y relativas (porcentajes) para intervalos de valores de amplitud 5, denotados por su punto medio. Esa misma distribución de frecuencias se ha representado gráficamente mediante el histograma que mostramos en la figura II.

TABLA I
Distribución muestral empírica para el estadístico $\bar{y}^ - \bar{z}^*$*

$\bar{y}^* - \bar{z}^*$	Frecuencia	%	% acumulado
-30	1	0,0	0,0
-25	9	0,2	0,2
-20	38	0,8	1,0
-15	186	3,7	4,7
-10	568	11,4	16,0
-5	1073	21,5	37,5
0	1256	25,1	62,9
5	1073	21,5	84,1
10	549	11,0	95,1
15	196	3,9	99,0
20	43	0,9	99,8
25	8	0,2	100,0

FIGURA II
Histograma para los valores observados del estadístico $\bar{y}^ - \bar{z}^*$.*



La distribución obtenida constituye una estimación por medio de Bootstrap de la distribución muestral para el estadístico diferencia de medias. Si contamos el número de veces que la diferencia de medias supera o iguala al valor observado en las muestras originales, que es de 12,1, podremos obtener la frecuencia relativa, que consideraremos como una aproximación a la probabilidad de encontrar, aceptando la hipótesis nula como cierta, una diferencia de medias igual o mayor a la observada:

$$\begin{aligned} \text{Prob}_{H_0} \{(\bar{y}^* - \bar{z}^*) \geq 12,1\} &= \\ &= \#\{(\bar{y}^* - \bar{z}^*) \geq 12,1\} / 5000 \end{aligned}$$

Estimada a partir de la distribución empírica, en la que 239 muestras han dado lugar a diferencias iguales o mayores que 12,1, el valor de esta probabilidad resulta ser $p=0,0478$.

Trabajando con un nivel de significación de 0,05, podríamos afirmar que existe una diferencia significativa a favor del

segundo grupo (contraste unilateral derecho), o lo que es igual, que las actitudes de los alumnos hacia la investigación cualitativa son mejores tras la experiencia realizada en el marco de la asignatura Métodos Cualitativos de Investigación Educativa.

CONTRASTE BASADO EN EL ESTADÍSTICO t

El segundo de los ejemplos que presentamos fue tomado de la experiencia realizada por Gil y García (2002), consistente en la utilización y valoración de una *Guía Interactiva sobre Investigación Educativa*, diseñada para orientar a los alumnos de Ciencias de la Educación sobre las posibilidades de realizar investigación educativa, tanto desde el contexto académico como profesional, y motivarles hacia la misma. Dicha experiencia se llevó a cabo con estudiantes de la Facultad de Ciencias de la Educación de la Universidad de Sevilla, que cursaban asignaturas de métodos de investigación educativa en el curso académico 2001-02.

TABLA II
*Distribuciones de frecuencias y estadísticos descriptivos
 para las valoraciones en cada grupo*

Grupo z		Grupo y	
Valor.	Frec.	Valor.	Frec.
1	1	1	0
2	5	2	1
3	9	3	8
4	6	4	14
5	3	5	4
$\bar{z} = 3,21;$		$\bar{y} = 3,78;$	
$S_z = 1,06$		$S_y = 0,75$	

Entre las técnicas utilizadas para obtener las opiniones de los usuarios, se recurrió a una escala de valoración, que los alumnos y alumnas completaron una vez examinado el material multimedia. Éstos otorgaron una puntuación comprendida entre 1 y 5 a diferentes aspectos, entre los que se encuentra el hecho de que la información contenida en la Guía constituya un estímulo para hacer investigación. Para determinar diferencias entre el poder motivador que se atribuye a la Guía en distintos colectivos destinatarios de la misma, consideraremos las valoraciones realizadas por 24 alumnos de 2º de Pedagogía (grupo z) y 27 alumnos de 1º de Psicopedagogía (grupo y). En la tabla II se incluyen las correspondientes distribuciones de frecuencias, así como los estadísticos media y desviación típica para las valoraciones obtenidas de ambos grupos.

Teniendo en cuenta que ambos grupos poseen experiencias académicas y expectativas diferentes, y a la vista de los estadísticos media que se alcanzan, podría sospecharse la existencia de formas diferenciadas de valorar la capacidad estimuladora del material multimedia diseñado. Para comprobarlo, someteremos a

contraste la hipótesis nula de igualdad de medias.

En este caso, no se verifica el supuesto de normalidad exigido para la aplicación de un contraste paramétrico basado en el estadístico t. La prueba de bondad de ajuste de Kolmogorov-Smirnov, utilizada para comprobar la hipótesis nula de que no hay diferencias significativas entre la distribución empírica y el modelo de distribución normal, arrojó para los grupos de alumnos de Pedagogía y Psicopedagogía grados de significación $p=0,277$ y $p=0,026$ respectivamente. En consecuencia, puede afirmarse para el segundo grupo que la muestra extraída no procede de una población distribuida normalmente, y hablando en términos estrictos, la distribución muestral del estadístico t no puede ser identificada con un modelo teórico de distribución.

De cara a hacer frente a esta situación, una alternativa no paramétrica se encuentra en la aplicación del método Bootstrap para el contraste de medias. Aunque podríamos seguir un procedimiento simple similar al mostrado en el primero de los ejemplos, que mostrábamos en el apartado anterior, aquí recurriremos al

estadístico t, ejemplificando el modo en que los procedimientos tradicionales pueden ser reenfocados desde la óptica de las técnicas de remuestreo. Un modo de actuar sería el que describimos seguidamente:

- Combinamos los datos disponibles, que son los obtenidos para los grupos z e y, obteniendo el conjunto $x = \{z, y\}$ que consta de un total de 51 valores. Extraemos por muestreo aleatorio con reposición una muestra x^* de tamaño $n=51$, constituyendo con los primeros valores una nueva muestra z^* de tamaño $p=24$ y con los restantes una muestra y^* de tamaño $m=27$.
- Para la muestra x^* , calculamos el valor del estadístico t, que vendrá dado por la expresión:

$$t(x^*) = \frac{\bar{y}^* - \bar{z}^*}{\sqrt{\frac{\sigma_y^{2*}}{m} + \frac{\sigma_z^{2*}}{p}}}$$

donde

$$\sigma_y^{2*} = \sum_1^m (y_i^* - \bar{y}^*)^2 / (m - 1);$$

$$\sigma_z^{2*} = \sum_1^p (z_i^* - \bar{z}^*)^2 / (p - 1)$$

- Hemos adoptado el estadístico t que se utiliza cuando no se asume el supuesto de igualdad de varianzas.
- Repetimos los pasos primero y segundo un número elevado de veces, que de nuevo en este caso será de 5.000.
- Construimos la distribución de frecuencias para los valores asumidos por el estadístico t (x^*). Esta distribución empírica y el histograma elaborado para su representación gráfica se muestran respectivamente en la tabla III y la figura III.

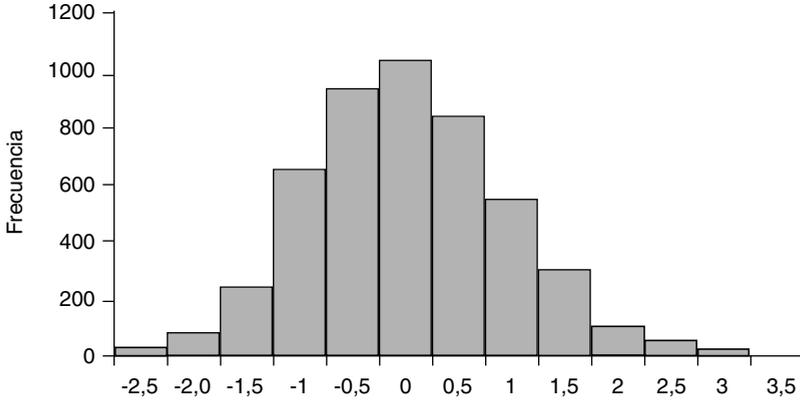
TABLA III

Distribución muestral empírica para el estadístico t (x^)*

T(x^*)	Frecuencia	%	% acumulado
-2,5	11	0,2	0,2
-2,0	103	2,1	2,3
-1,5	267	5,3	7,6
-1,0	688	13,8	21,4
-0,5	950	19,0	40,4
0,0	1040	20,8	61,2
0,5	859	17,2	78,4
1,0	566	11,3	89,7
1,5	310	6,2	95,9
2,0	131	2,6	98,5
2,5	57	1,1	99,6
3,0	17	0,3	100,0
3,5	1	0,0	100,0

FIGURA II

Histograma para los valores observados del estadístico $t(x^)$*



Dado que la ausencia de normalidad impide afirmar que la distribución muestral del estadístico t sigue el modelo de la t de Student, tomamos la distribución empírica como base para estimar la probabilidad de encontrar valores iguales o mayores que el valor observado de t (contraste unilateral derecho), que calculado para las muestras originales es $t=2,185$. Puesto que sólo se alcanza o supera este valor en 83 de un total de 5000 muestras utilizadas, la probabilidad que buscamos resulta ser:

$$\begin{aligned} \text{Prob}_{H_0} \{ (t(x^*) \geq 2,185) \} &= \\ &= \# \{ (t(x^*) \geq 2,185) \} / 5000 = 0,0166 \end{aligned}$$

Para un nivel de significación de 0,05, podríamos rechazar la hipótesis nula y afirmar que existen diferencias significativas a favor del segundo grupo, por lo que la capacidad motivadora atribuida a la *Guía Interactiva* por los alumnos de Psicopedagogía es mayor que entre los alumnos de Pedagogía.

VALORACIÓN Y CONCLUSIONES

Un primer aspecto a valorar es la utilidad de los métodos Bootstrap en las situaciones en las que queda comprometida la idoneidad de los enfoques paramétricos clásicos para el contraste de hipótesis. Tanto unos como otros implican partir de la muestra observada con el propósito de llegar a conclusiones sobre una población, pero mientras que la estadística paramétrica se apoya en supuestos sobre la distribución poblacional o sobre sus parámetros, el Bootstrap prescinde de ellos. La corrección de los resultados obtenidos por medio de Bootstrap para la inferencia estadística ha sido demostrada a partir de su utilización en simulaciones, en las que se analizaban muestras extraídas de poblaciones distribuidas normalmente. En tales circunstancias, la aplicación de los métodos paramétricos y métodos Bootstrap dio lugar a resultados similares en la estimación de medias o en la estimación de coeficientes para la regresión (Mooney y Duval, 1993).

Si bien esta comparación permite incrementar la confianza sobre los métodos Bootstrap, bien es verdad que si los supuestos en los que se basa la descripción matemática de la población se cumplen, no hay razones para descartar los procedimientos de inferencia estadística que posibilita el enfoque tradicional. La utilidad de este tipo de métodos se hace patente, de manera especial, en aquellas situaciones en las que no es posible justificar los supuestos de partida o cuando no se cuenta con fórmulas analíticas en las que basar nuestra descripción de la población.

En la práctica del análisis estadístico, la alternativa habitual a las pruebas paramétricas para el contraste de hipótesis se ha apoyado en procedimientos que tienen como ventaja el no hacer supuestos sobre las poblaciones o sobre sus parámetros, y que se han englobado bajo la denominación de estadística no paramétrica (Siegel, 1991). Sin embargo, la mayor parte de estas técnicas suponen trabajar con variables a nivel ordinal, reduciendo a rangos las puntuaciones observadas y perdiendo por tanto una parte de la información disponible. Este problema queda soslayado en los métodos Bootstrap, los cuales permiten en tales situaciones conservar el nivel de medida incluso de variables en escala de intervalos.

Aplicadas en los dos ejemplos que hemos presentado, las pruebas no paramétricas para el contraste de grupos conducirían a resultados inferiores a los obtenidos aquí mediante Bootstrap. Así en el primer ejemplo, la aplicación de la prueba de Mann-Whitney para dos muestras independientes arroja para el estadístico de contraste un valor $U=6,50$, al que corresponde en el caso de una prueba unilateral la probabilidad $p=0,060$. En el segundo de los ejemplos, utilizando la misma prueba no paramétrica obtenemos

$U= 218,50$, con un grado de significación $p=0,018$ para el contraste unilateral. En consecuencia, comparando con los valores de p estimados mediante Bootstrap ($0,0478$ y $0,0166$ respectivamente), observamos en éstos últimos una mejora del error Tipo I cometido, es decir, se reduce la probabilidad de rechazar una hipótesis nula que sea cierta. Gracias a ello, ha sido posible en el primer caso rechazar H_0 con un nivel de significación $\alpha=0,05$ llegando a un resultado no alcanzado mediante la prueba U de Mann-Whitney.

Otro aspecto destacable es la simplicidad con la que puede aplicarse el método Bootstrap. Un ejemplo de ello es la primera de las aplicaciones que aquí hemos presentado, en el cual utilizamos únicamente la diferencia de medias y no precisamos de ningún tipo de expresión algebraica para describir los parámetros de su distribución muestral. Esta simplicidad hace que el método constituya un enfoque atractivo en la enseñanza de la estadística. La simulación a partir de muestras permite trabajar sin fórmulas ni descripciones matemáticas, que no siempre son comprendidas por el alumnado y que a menudo constituyen un obstáculo para el aprendizaje.

Un inconveniente de los métodos estadísticos convencionales está no sólo en el manejo correcto de las nociones aritméticas, sino también en la elección correcta de las fórmulas que es preciso aplicar en cada situación. Para algunos estudiantes, el manejo de las fórmulas y métodos estadísticos llega a adquirir un carácter mágico. Saliendo al paso de esta situación, los métodos basados en el remuestreo presentan como ventaja la utilización de técnicas simples e intuitivas basadas en la simulación de un modelo a partir de un número elevado de muestras aleatorias. La resolución de los problemas estadísticos deja de estar vinculada a la pericia matemática y pasa a ser una

cuestión de claridad de pensamiento sobre los problemas planteados.

A pesar del peso de tales argumentos, los métodos Bootstrap, al igual que otros desarrollos estadísticos de las últimas décadas, siguen estando ausentes de la mayor parte de los textos introductorios a las técnicas estadísticas (Efron, 2000, p. 1295). La introducción en el currículum de este tipo de técnicas, que permiten un acercamiento intuitivo a la estadística, es uno de los retos que siguen pendientes en relación con el desarrollo del Bootstrap.

La viabilidad de poner en práctica enfoques como el Bootstrap se basa en el avance de los ordenadores, cada vez más accesibles, fáciles de usar, y rápidos en la realización de cálculos. La evolución de este tipo de métodos estadísticos basados en la computación intensiva presenta, a juicio de autores como Chernick (1999), un brillante futuro. Cabe esperar que en el horizonte inmediato de la investigación educativa, la utilización del Bootstrap vaya cobrando importancia y la comunidad científica de nuestro ámbito reconozca las posibilidades de esta nueva forma de inferencia estadística, de tal manera que se alcancen niveles de aplicación similares a los que ya se están dando en otras disciplinas.

Antes de terminar, es preciso señalar también alguna de las limitaciones del Bootstrap, derivada de la propia idea que le sirve de base. Si el método se apoya en la analogía entre la muestra observada y la población de la que fue extraída, la calidad de la muestra es crucial. En este sentido, los resultados del Bootstrap se ven afectados cuando la muestra no se extrae por un procedimiento de muestreo aleatorio simple y cuando el tamaño de ésta es demasiado pequeño (Mooney y Duval, 1993). En ambas situaciones, puede cuestionarse que la información ofrecida por la muestra permita reconstruir

adecuadamente la población estudiada. Sin embargo, esta limitación podría ser trasladable a otras técnicas paramétricas y no paramétricas, las cuales pierden potencia al reducirse el tamaño de las muestras. Ante esta circunstancia, cabría afirmar que los métodos Bootstrap permiten «extraer lo máximo a partir de la poca información disponible» (Chernick, 1999, p. 149).

BIBLIOGRAFÍA

- CHANG, L.: «Quantitative Attitudes Questionnaire: Instrument development and validation», en *Educational and Psychological Measurement*, 56, 6 (1996), pp. 1037-1042.
- CHEARNICK, M. R.: *Bootstrap methods: a practitioner's guide*. Nueva York, Wiley & Sons, 1999.
- DAVISON, A.; HINKLEY, D.: *Bootstrap methods and their application*. Nueva York, Cambridge University Press, 1997.
- EFRON, B.: «Bootstrap methods: another look at the jackknife», en *The Annals of Statistics*, 7 (1979), pp. 1-26.
- *The jackknife, the Bootstrap and other resampling plans*. Philadelphia, Society for Industrial and Applied Mathematics, 1982.
- «The statistical century», en *RSS News*, 22, 5 (1995), pp. 1-2.
- «The Bootstrap and modern statistics», en *Journal of the American Statistical Association*, 95, 452 (2000), pp. 1293-1296.
- EFRON, B.; TIBSHIRANI, R. J.: *An introduction to the Bootstrap*. Nueva York, Chapman & Hall/CRC, 1993.
- FAN, X.: «Using Commonly Available Software for Conducting Bootstrap Analyses». Comunicación presentada al *Annual Meeting of the American Educational Research Association*. Seattle, WA, 2001.

- GIL, J.; GARCÍA, S.: *Motivación de los estudiantes de Ciencias de la Educación hacia la investigación en el ámbito educativo*. Informe inédito, 2002.
- GIL, J.; JAÉN, A.: «La realización de proyectos como eje de la estrategia docente en la enseñanza de los métodos de investigación educativa», en *Revista de Enseñanza Universitaria* (en prensa).
- GUTHRIE, A.: «Using Bootstrap Methods with Popular Statistical Programs». Comunicación presentada al *Annual Meeting of the Southwest Educational Research Association*. Nueva Orleans, 2001.
- HALL, P.: *The Bootstrap and edgeworth expansion*. Nueva York, Springer, 1992.
- HANCOCK, G. R.; NEVITT, J.: «Bootstrapping and the Identification of Exogenous Latent Variables within Structural Equation Models», en *Structural-Equation-Modeling*, 6, 4 (1999), pp. 394-99.
- HARRIS, D. J.; KOLEN, M. J.: «Bootstrap and Traditional Standard Errors of the Point Biserial», en *Educational and Psychological Measurement*, 48, 1 (1998), pp. 43-51.
- ICHIKAWA, M.; KONISHI, S.: «Application of the Bootstrap Methods in Factor Analysis», en *Psychometrika*, 60, 1 (1995), pp. 77-93.
- KNAPP, T. R.; NOBLITT, G. L.; VIRAGOONTAVAN, S.: «Traditional vs. "Resampling" Approaches to Statistical Inferences Regarding Correlation Coefficients», en *Mid-Western Educational Researcher*, 13, 2 (2000), pp. 34-36.
- MILLER, R. G.: «The jackknife: a review», en *Biometrika*, 61 (1974), pp. 1-17.
- MOONEY, C. Z.: *Monte Carlo simulation*. Thousand Oaks, Sage Publications, 1997.
- MOONEY, C. Z.; DUVAL, R. D.: *Bootstrapping. A nonparametric approach to statistical inference*. Newbury Park, Sage Publications, 1993.
- NEVITT, J.; HANCOCK, G. R.: «Performance of Bootstrapping Approaches To Model Test Statistics and Parameter Standard Error Estimation in Structural Equation Modeling», en *Structural Equation Modeling*, 8, 3 (2001), pp. 353-377.
- NOREEN, E.: *Computerintensive methods for testing hypotheses*. New York, Wiley, 1989.
- QUENOUILLE, M. H.: «Approximate tests of correlation in time series», en *Journal of the Royal Statistical Society*, 11 (1949), pp. 18-84.
- RAYKOV, T.: «Approximate Confidence Interval for Difference of Fit in Structural Equation Models», en *Structural Equation Modeling*, 8, 3 (2001), pp. 458-469.
- RESAMPLING STATS INC.: *Resampling Stats Add-In for Excel, version 2.0*, 2001.
- RODGERS, J. L.: «The Bootstrap, the jackknife, and the randomization test: a sampling taxonomy», en *Multivariate Behavioral Research*, 34, 4 (1999), pp. 441-456.
- ROMO, J.: «Técnicas Bootstrap en econometría: una introducción», en *Cuadernos Económicos del ICE*, 56 (1994), pp. 179-194.
- SHAO, J.; TU, D.: *The jackknife and the Bootstrap*. Nueva York, Springer, 1995.
- SIEGEL, S.: *Estadística no paramétrica*. México, Trillas, 1991.
- SIMON, J. L.: *Basic research methods in social science*. New York, Random House, 1969.
- TUKEY, J. W.: «Bias and confidence in not quite large samples», en *American Mathematical Statistics*, 29 (1958), p. 614.
- YOUNG, G. A.: «Bootstrap: more than a stab in the dark?», en *Statistics Sciences*, 9 (1994), pp. 382-415.