ALTERNATIVAS AL ANÁLISIS DE LA VARIANZA MEDIANTE TÉCNICAS BOOTSTRAP

Javier Gil Flores
Universidad de Sevilla

Las técnicas bootstrap

La inferencia estadística convencional se apoya en buena medida en el supuesto de que trabajamos con variables que se distribuyen normalmente en las poblaciones objeto de estudio. Partiendo de éste y otros supuestos acerca de los parámetros poblacionales, somos capaces de determinar la distribución muestral teórica de determinados estadísticos, que es utilizada para hacer inferencias sobre parámetros poblacionales o para el contraste de hipótesis planteadas acerca de los mismos.

Las técnicas bootstrap representan un modo diferente de afrontar el problema de la inferencia estadística. El punto de partida sigue siendo la información proporcionada por la muestra, pero en lugar de usar ésta para formular modelos teóricos de distribución atribuidos a los estadísticos muestrales en la población de referencia, se recurre a procedimientos de remuestreo que permiten generar a partir de ella un número elevado de nuevas muestras. Basándose en este amplio conjunto de muestras es posible construir la distribución muestral empírica de los estadísticos objeto de interés, y en función de ella realizar inferencias. El bootstrap se incluye por tanto entre los procedimientos de remuestreo, que tienen en común la obtención de muestras a partir de los valores observados con el fin de generar distribuciones muestrales empíricas. En el caso del bootstrap, las muestras generadas tienen el mismo tamaño que la muestra observada, y el muestreo se realiza con reposición.

Aunque las ideas que les sirven de base habían sido consideradas en momentos anteriores, la primera formalización de las técnicas bootstrap se debe a Efron (1979). La utilización del bootstrap en el análisis estadístico ha sido abordada con amplitud en textos y artículos que han visto la luz con posterioridad, como los de Bickel y Friedman (1981), Singh (1981); Romano (1988), Hall (1992), Efron y Tibshirani (1993), Shao y Tu (1995), Davison y Hinkley (1997) o Chernick (1999). El bootstrap no constituye un procedimiento estadístico particular, sino un enfoque amplio que significa otro modo de aplicar la estadística para hacer inferencias. De hecho, su utilización en la estimación de parámetros, los contrastes para dos muestras, el análisis de la varianza, la regresión, las series temporales, el análisis factorial y en otros métodos de análisis multivariante suele ser presentada en los manuales y estudios dedicados a este tema, mostrando el amplio alcance que tiene la aplicación de las técnicas bootstrap.

La utilización del bootstrap ha logrado cierta presencia en una variedad de disciplinas tan dispares como la psicología, la física, la econometría, la medicina o la ingeniería (Chernick, 1999:7), aunque en el caso de la educación su desarrollo ha sido algo menor hasta el momento. De hecho, en nuestro contexto más próximo, las técnicas bootstrap están ausentes de los análisis estadísticos realizados por los investigadores educativos.

Aplicación del bootstrap en el ANOVA

En el presente trabajo, mostraremos las posibilidades que ofrece el bootstrap en el contexto del análisis de la varianza. Para ello recurriremos a datos extraídos de un estudio realizado para valorar las actitudes hacia la Estadística entre estudiantes de Pedagogía y la incidencia en ellas de variables como la formación previa (Gil, 1999). Concretamente, en este estudio se aplicaba la Escala de Actitudes hacia la Estadística (Wise, 1985), que consta de 29 items ante los cuales el alumnado expresa su grado de acuerdo o desacuerdo utilizando para ello una escala de cinco puntos. El análisis de los datos obtenidos tras la aplicación de este instrumento a una amplia muestra de estudiantes apuntaba la existencia de diversos factores, entre los que se encontraba el factor *ansiedad ante la Estadística*. A efectos de nuestra reflexión metodológica sobre las posibilidades de *bootstrapping* en el análisis de la varianza, aquí tomaremos las puntuaciones obtenidas en este factor por tres pequeñas muestras de 21, 21 y 16 alumnos que habían cursado respectivamente bachilleratos de Ciencias, Letras y otras modalidades de bachillerato antes de su ingreso en la universidad. Los estadísticos muestrales para los tres grupos de alumnos aparecen en la tabla 1.

Tabla 1: Medias y desviaciones típicas para los tres grupos.

Grupo	*n*	*Media*	*Desv. Típica*
Bachillerato Ciencias	21	3,652	0,619
Bachillerato Letras	21	2,952	0,586
Otros Bachilleratos	16	3,375	0,748

Trabajando con estos datos, nos cuestionaremos la existencia de diferencias en la *ansiedad ante la Estadística* entre los tres grupos de alumnos considerados. El método de análisis habitual para responder a esta cuestión es el análisis de la varianza de un factor, que permitiría contrastar la hipótesis nula de que no existen diferencias significativas entre las medias poblacionales para alumnos procedentes de distintos bachilleratos. La aplicación de esta prueba estadística exige la comprobación previa de ciertos supuestos, que se concretan en la normalidad de las poblaciones y la homoscedasticidad de varianzas. Para comprobar la normalidad poblacional, se ha recurrido a la prueba de Kolmogorov-Smirnov (ver tabla 2), cuyos resultados permiten mantener que la variable *ansiedad ante la Estadística* se distribuye normalmente en las tres poblaciones consideradas. En el caso de la homoscedasticidad de varianzas se recurrió a la prueba de Levene, cuya aplicación arrojó un valor F=0,803, al que corresponde un grado de significación p=0,453. De acuerdo con este valor, también podemos mantener la hipótesis nula de igualdad de varianzas.

Tabla 2: Prueba de Kolmogorov-Smirnov para el contraste de la normalidad poblacional

Grupo	*Diferencias más extremas*			*Z*	*P*
	Absoluta	*Positiva*	*Negativa*		
Bachillerato Ciencias	0,145	0,086	-0,145	0,665	0,769
Bachillerato Letras	0,187	0,091	-0,187	0,859	0,452
Otros Bachilleratos	0,215	0,135	-0,215	0,860	0,450

Comprobados los supuestos paramétricos, procederíamos a calcular el estadístico F que compara la variabilidad intergrupos o experimental con la variabilidad intragrupos o de error (ver tabla 3). El valor obtenido es F=6,2424. Considerando que el estadístico F sigue el modelo teórico de la distribución F para 2 y 55 grados de libertad, al valor observado corresponde un grado de significación p=0,0036 de acuerdo con el cual podríamos rechazar la hipótesis nula de igualdad de medias y afirmar, para un nivel de significación a=0,01, la existencia de distintos niveles de ansiedad ante la Estadística entre los alumnos que proceden de diferentes modalidades de Bachillerato.

Tabla 3: Análisis de la varianza para la variable ansiedad ante la Estadística

	Suma de cuadrados	*g.l.*	*Media cuadrática*	*F*	*Sig.*
Inter.-grupos	5,206	2	2,603	6,2424	0,0036
Intra-grupos	22,935	55	0,417		
Total	28,141	57			

Hasta aquí hemos mostrado el modo de proceder en el análisis de la varianza, obteniendo finalmente un valor p en función del cual adoptar la decisión estadística. No obstante, podríamos obtener también este valor p mediante *bootstrapping*, basándonos para ello en diferentes estadísticos. En tal caso, no sería necesario verificar ningún tipo de supuestos de partida acerca de las distribuciones o los parámetros poblacionales.

Para llevar a cabo el análisis hemos utilizado el software *Resampling Stats Add-In for Excel* (Resampling Stats Inc., 2001). El modo de proceder sería el siguiente:

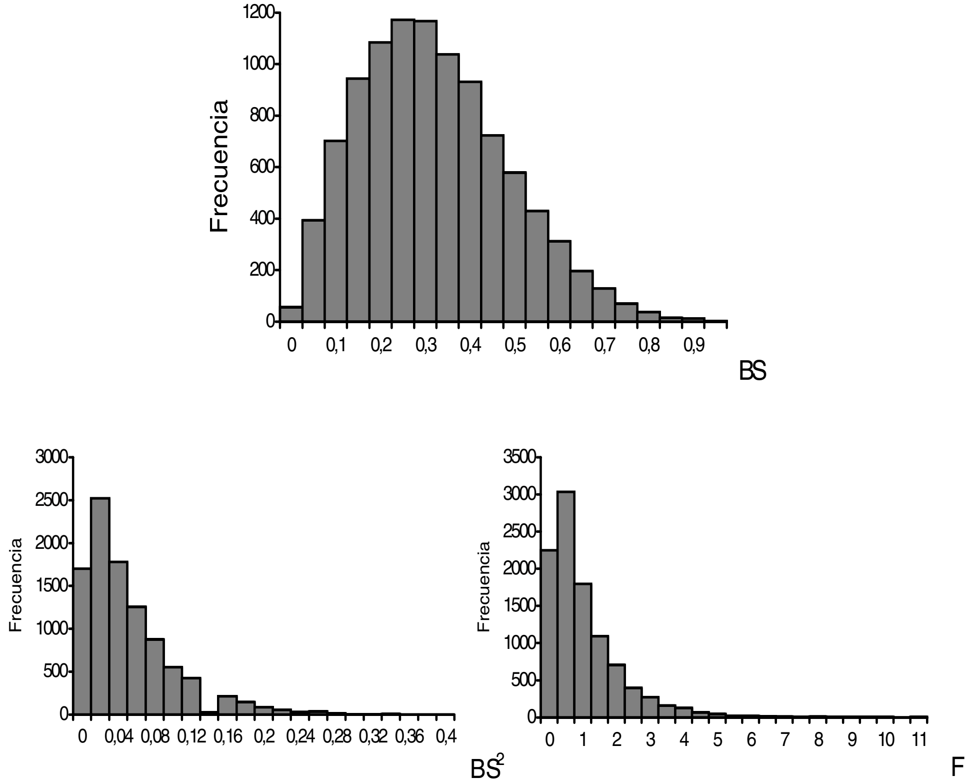
1. La hipótesis nula sometida a contraste establece la igualdad de los tres grupos de alumnos en cuanto a la ansiedad ante la estadística. Si H_0 es cierta, las puntuaciones alcanzadas en las tres muestras proceden de una misma población, por lo que el conjunto de todas ellas constituye la información disponible sobre la población de referencia. Usando las 58 puntuaciones, podemos extraer con reposición una nueva muestra de tamaño n=58, asignando los 21 primeros valores al primer grupo, los 21 siguientes al segundo y los 16 restantes al tercero y último.
2. La muestra obtenida según el procedimiento descrito permite calcular el valor de un estadístico de contraste. Siguiendo a Simon y Bruce (1991), podríamos utilizar la suma de las diferencias absolutas de cada media grupal respecto a la media global (SB); otra posibilidad estaría en la utilización de la suma de las diferencias cuadráticas (SB^2); y finalmente, cabría utilizar el propio estadístico de contraste calculado en el análisis de la varianza (F). Los dos primeros estadísticos aportan una medida del grado en que los grupos se diferencian entre sí, esperándose valores próximos a cero bajo la hipótesis nula de igualdad entre los mismos. Aunque bastaría emplear cualquiera de los estadísticos mencionados, aquí utilizaremos los tres para llevar a cabo el contraste y mostrar de este modo las diferentes alternativas a la prueba clásica.
3. Los dos pasos anteriores se repiten hasta conseguir un número elevado de muestras, en cada una de las cuales se pueden calcular cualquiera de los estadísticos de contraste. Si bien 5.000 muestras podrían resultar suficientes (Manly, 1997;Chernick, 1999), trabajaremos sobre 10.000 muestras considerando que el aumento en el número de éstas mejorará la estimación de la distribución muestral para los estadísticos utilizados.
4. Los valores observados para los estadísticos de contraste en las muestras extraídas permiten construir la distribución muestral empírica para tales estadísticos. En la tabla 4 se muestran las distribuciones de frecuencias para los valores observados en el caso de los estadísticos SB, SB^2 y F. En

las tres distribuciones los valores han sido agrupados por intervalos, denotando éstos por su punto medio. Las distribuciones obtenidas también han sido representadas gráficamente mediante histogramas (ver figura 1).

Tabla 4: Distribuciones muestrales empíricas para los estadísticos SB, SB^2 y F obtenidas mediante bootstrap.

BS	*Frec.*	% *acum..*	*BS^2*	*Frec.*	% *acum..*	*F*	*Frec.*	% *acum..*
0,00	57	0,6	0,00	1702	17,0	0,00	2242	22,4
0,05	394	4,5	0,02	2522	42,2	0,50	3028	52,7
0,10	701	11,5	0,04	1778	60,0	1,00	1793	70,7
0,15	943	21,0	0,06	1255	72,6	1,50	1087	81,5
0,20	1085	31,8	0,08	875	81,3	2,00	700	88,5
0,25	1172	43,5	0,10	554	86,9	2,50	395	92,5
0,30	1167	55,2	0,12	425	91,1	3,00	273	95,2
0,35	1038	65,6	0,14	284	93,9	3,50	161	96,8
0,40	932	74,9	0,16	215	96,1	4,00	124	98,1
0,45	724	82,1	0,18	148	97,6	4,50	65	98,7
0,50	580	87,9	0,20	87	98,4	5,00	43	99,1
0,55	429	92,2	0,22	55	99,0	5,50	21	99,3
0,60	313	95,3	0,24	32	99,3	6,00	21	99,6
0,65	197	97,3	0,26	35	99,7	6,50	14	99,7
0,70	129	98,6	0,28	12	99,8	7,00	8	99,8
0,75	71	99,3	0,30	4	99,8	7,50	5	99,8
0,80	38	99,7	0,32	5	99,9	8,00	6	99,9
0,85	15	99,8	0,34	9	100,0	8,50	3	99,9
0,90	12	100,0	0,36	2	100,0	9,00	3	99,9
0,95	3	100,0	0,40	1	100,0	9,50	1	100,0
						10,00	1	100,0
						11,00	4	100,0

Figura 1: Histogramas para los valores observados de los estadísticos BS, BS^2 y F.



Las distribuciones obtenidas mediante bootstrap constituyen estimaciones de las distribuciones muestrales correspondientes. Contando el número de veces que el valor del estadístico de contraste iguala o supera al valor observado en las muestras originales (BS=0,7781; BS^2=0,26771; F=6,2424), tendremos una aproximación a la probabilidad de encontrar valores más extremos siendo H_0 cierta. Por tanto, el valor de esta probabilidad representa el riesgo de rechazar una hipótesis nula verdadera.

Para cualquiera de los tres estadísticos utilizados, la probabilidad de encontrar valores más extremos resulta ser muy baja. Concretamente, hemos obtenido valores p=0.0076, p=0.0037 y p=0.0045 para cada uno de los respectivos estadísticos SB, SB^2 y F. Todas las vías utilizadas permitirían por tanto rechazar la hipótesis nula con un nivel de confianza superior al 99% (a=0.01), conduciendo a resultados muy similares a los obtenidos en el análisis de la varianza convencional.

CONCLUSIONES

La alta coincidencia entre los resultados obtenidos en la aplicación de técnicas bootstrap y el análisis de varianza corroboran la adecuación de este tipo de enfoques para la inferencia estadística. La superioridad de los mismos se hará evidente en situaciones en las que no se satisfacen los supuestos que permitirían la aplicación del análisis de la varianza. En tal caso, el recurso a técnicas no paramétricas implicaría trabajar con puntuaciones previamente transformadas en rangos, con la consiguiente pérdida de información. Mediante bootstrap, es posible conservar el nivel de medida y llevar a cabo las pruebas de contraste utilizando estadísticos cuyas distribuciones muestrales se obtienen empíricamente a partir del remuestreo sobre las muestras observadas.

Este tipo de enfoques presenta además como ventaja la simplicidad con que pueden ser aplicados. La dificultad y complejidad que rodea a veces al análisis estadístico constituye un obstáculo para muchos de quienes se inician en su estudio. El empleo de técnicas bootstrap permite obviar descripciones analíticas y formulaciones complejas para establecer las distribuciones muestrales que siguen los estadísticos. En el ejemplo aquí mostrado, bastaría trabajar con las diferencias de medias, sin necesidad de recurrir a expresiones algebraicas que a menudo resultan un escollo para el aprendizaje de la estadística por parte de los estudiantes.

Como limitación de estos enfoques habría que señalar la necesidad de contar con muestras los más representativas posibles. Puesto que el modo de operar consiste en reconstruir una población a partir de la muestra observada, resulta crucial la calidad de esta muestra. Además de la conveniencia de aleatoriedad en la extracción de la muestra, una consideración básica tendrá que ver con el tamaño, aspecto en el que las recomendaciones apuntan hacia un mínimo de entre 10 y 20 casos.

Uno de los motores para el desarrollo de las técnicas bootstrap es sin duda el avance logrado por la informática. La generación de un elevado número de muestras requerida para construir la distribución muestral empírica hace que este tipo de métodos sólo sea viable contando con el recurso de los ordenadores. Hasta el desarrollo de éstos, posiblemente la única forma de trabajar consistía en apoyar la inferencia estadística sobre modelos teóricos de distribución muestral atribuidos a los estadísticos, siempre que se cumplan ciertos supuestos. En palabras de Rodgers (1999:441), la forma en que realizamos los análisis estadísticos puede ser en parte fruto de un accidente histórico; si el análisis de la varianza de Fisher hubiera sido inventado treinta años más tarde (o los ordenadores hubieran estado disponibles treinta años antes), nuestros procedimientos estadísticos estarían probablemente menos vinculados a distribuciones teóricas como F o t.

Cabe esperar que la popularidad que estos métodos empiezan a tener en algunas disciplinas científicas, se extienda a otros campos y las técnicas bootstrap acaben ocupando un lugar destacado en el ámbito del análisis estadístico. Para Efron (2000), el bootstrap se encuentra entre los avances recientes que han tenido un mayor efecto sobre la práctica del análisis estadístico, y sin embargo no aparece en la mayor parte de los manuales sobre estadística. Su inclusión en el ámbito de la enseñanza de esta disciplina es una cuestión pendiente.

Referencias bibliográficas

BICKEL, P. y FRIEDMAN, D. (1981). Some asymptotic theory for the bootstrap, en *The Annals of Statistics*, 9, 1196-1217.

CHERNICK, M.R. (1999). *Bootstrap methods: a practitioner's guide*. Nueva York: Wiley & Sons.

DAVISON, A. y HINKLEY, D. (1997). *Bootstrap methods and their application*. Nueva York: Cambridge University Press.

EFRON, B. (1979). Bootstrap methods: another look at the jacknife, en *The Annals of Statistics*, 7, 1-26.

EFRON, B. (2000). The bootstrap and modern statistics, en *Journal of the American Statistical Association*, 95 (452), 1293-1296.

EFRON, B. y TIBSHIRANI, R.J. (1993). *An introduction to the bootstrap*. Nueva York: Chapman & Hall/CRC.

GIL, J. (1999). Actitudes hacia la estadística. Incidencia de las variables sexo y formación previa, en *Revista Española de Pedagogía*, LVII, 214, 567-590.

HALL, P. (1992). *The bootstrap and edgeworth expansion*. Nueva York: Springer.

MANLY, B.F. (1997). Randomization, bootstrap and Monte Carlo methods in Biology. Londres: Chapman & Hall.

RESAMPLING STATS INC. (2001). *Resampling Stats Add-In for Excel, version 2.0*.

RODGERS, J.L. (1999). The bootstrap, the jacknife, and the randomization test: a sampling taxonomy, en *Multivariate Behavioral Research*, 34 (4), 441-456.

ROMANO, J. (1988). A bootstrap revival of some nonparametric distance tests, en *Journal of the American Statistical Association*, 83, 698-708.

SHAO, J. y TU, D. (1995). *The jacknife and the bootstrap*. Nueva York: Springer.

SIMON, J.L. y BRUCE, P. (1991). Resampling: a tool for everiday statistical work, en *Chance*, 4 (1), 22-32.

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap, en *The Annals of Satatistics*, 9, 382-415.

WISE, S. (1985). The development and validation of a scale measuring attitudes toward statistics, en *Educational and Psychological Measurement*, 45, 401-405.