

The bootstrap: a 35 years old young very useful for analyzing biological data¹

El bootstrap: un joven de 35 años muy útil para analizar datos biológicos

J. A. Navarro Alberto

Departamento de Ecología Tropical

Campus de Ciencias Biológicas y Agropecuarias

Universidad Autónoma de Yucatán

Km 15.5 Carretera Mérida-Xmatkuil. CP 97315. Mérida, Yucatán, México.

Email: jorge.navarro@correo.uady.mx

In diverse scientific discussion forums and in specialized journals the word “bootstrap” has been mentioned or read. Will it be that for analyzing our data we must use boots with straps for its use as support points for jumping? (How odd...but this is the word for word translation!). In this paper the significance and development of the bootstrap is reviewed as rigorous statistical calculation method for data analysis. The diverse algorithms associated with the estimation by bootstrap intervals are shown and its application with the problem relative to mean estimation is explained. Finally mention is made of the implications and limitations of this method, as well as of the great usefulness it has in biological and agricultural sciences which have adopted it as analysis tool since its invention by Bradley Efron (U. of Stanford) since more than 35 years ago.

Key words: Bootstrap, interval of confidence, biological data, median

Introduction

In statistics it is customary to discuss about means and errors (standard) of variables in continuous scale: weights of seven months old steers, daily milk yield in dairy cows, etc. The strategy for the analysis of these variables for statistical inference is assuming that there is only statistical variation and that having a sufficient number of measurements is enough so as to calculate the mean and the standard error of the mean, knowing that the standard error decreases as the number of measurements increases. The artillery of statistical methods in these cases is vast and it is commanded by one of the most important theorems of Statistics: the Theorem of Central Limit: “If random size samples are taken n , y_1, y_2, \dots, y_n of a population with finite mean μ and variance σ^2 , then for an n sufficiently large, the sampling mean distribution (sampling means) can be approximated with a function of normal density with mean $\mu_{\bar{y}} = \mu$ and standard deviation (standard error) $\sigma_{\bar{y}} = \sigma/\sqrt{n}$. The error that will be committed due to the approximation of the distribution of the sampling means to the normal will decrease as n increases. For example,

En diversos foros de discusión de científicos, en revistas especializadas hemos oído mencionar o leído la palabra “bootstrap”. ¿Será que para analizar nuestros datos debemos usar botas con cintillos y usar éstos como puntos de apoyo para saltar? (¡Que cosa más rara... pero esta es la traducción literal de la palabra!). En este trabajo revisaremos el significado y desarrollo del bootstrap como método estadístico de cómputo intensivo para el análisis de datos. Se mostrarán los diversos algoritmos asociados con la estimación por intervalos bootstrap y se ilustrará su aplicación con el problema relativo a la estimación de la mediana. Finalmente, se hará mención de los alcances y limitaciones de este método, así como la gran utilidad que tiene dentro de las ciencias biológicas y agropecuarias, que las han adoptado como herramienta de análisis desde su invención por Bradley Efron (U. de Stanford) hace más de 35 años.

Palabras clave: Bootstrap, intervalos de confianza, datos biológicos, mediana

Introducción

En la estadística clásica estamos acostumbrados a hablar de promedios y errores (estándar) de variables en escala continua tales como: pesos de novillos de siete meses de edad, rendimiento diario de leche en vacas lecheras. La estrategia de análisis de estas variables para fines de inferencia estadística es asumir que solamente existe variación estadística y que basta con tener un número de mediciones suficiente para que podamos calcular la media y el error estándar de la media, sabiendo que el error estándar decrece conforme el número de mediciones crece. La artillería de métodos estadísticos en estos casos es vasta, siendo comandada por uno de los teoremas más importantes de la Estadística: el Teorema del Límite Central: “Si se toman muestras aleatorias de tamaño n , y_1, y_2, \dots, y_n de una población con media finita μ y varianza σ^2 , entonces para n suficientemente grande, la distribución de muestreo de medias (las medias muestrales) se pueden aproximar con una función de densidad normal con media $\mu_{\bar{y}} = \mu$ y desviación estándar (error estándar) $\sigma_{\bar{y}} = \sigma/\sqrt{n}$. El error que se cometerá por aproximar la distribución de muestreo de medias a la normal disminuirá conforme n aumente. Por ejemplo, en la figura 1, las distribuciones de

¹Paper presented at the V Congreso de Producción Animal Tropical, La Habana, Cuba, 2015

in figure 1, the sampling distributions generated on selecting the samples of normal random variables or uniform are approximately normal even for n as small as 10 (figures A1, B1). On the other hand, for biased distributions such as square ji the mean sampling distribution is not normal for n = 10 (figure C1), and only for n values as large as 100 is that this distribution is practically normal (figure C2).

In many cases, our variables of interest have unknown distribution and thus, it is not possible to know how large the sample must be in order to apply the result of the Theorem of Central Limit. If we rely on the results (asymptotic) of the Theorem of Central

muestreo generadas seleccionando muestras de variables aleatorias normales o uniformes son aproximadamente normales aun para n tan pequeñas como 10 (figuras A1, B1). En cambio, para distribuciones sesgadas como la ji cuadrada la distribución de muestreo de medias se observa no normalidad para n=10 (figura C1) y sólo para valores de n tan grandes como 100 es que tal distribución es prácticamente normal (figura C2).

En muchos casos, nuestras variables de interés tienen distribución desconocida y, por ende, no es posible saber cuán grande debe ser la muestra para poder aplicar el resultado del Teorema del Límite Central. Si uno se apoyara de los resultados (asintóticos) del

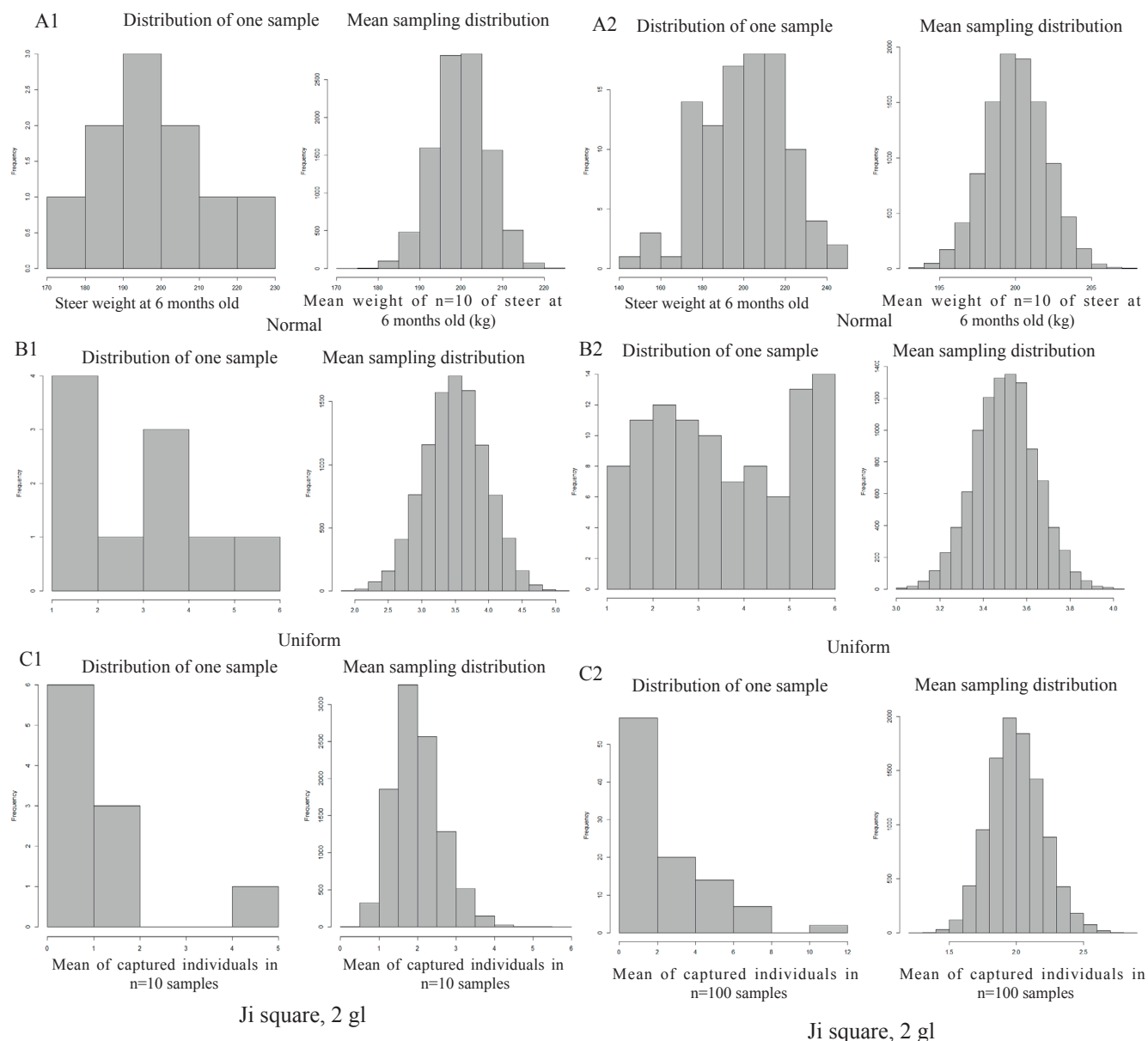


Figure 1. Distributions of sampling means generated on selecting samples of normal random variables (A1, A2) uniform (B1, B2) or square ji with 2 degrees of freedom (C1, C2). Left figures in each panel display the distribution of the frequencies observed for one of the samples of size n selected randomly of each distribution. For producing the distributions of mean sampling of each variable 1000 samples were generated of sizes n = 10 and n = 100. The random normal and uniform variables generated sampling distributions that can be approximated with the normal, regardless of the size of each sample. On the other hand, for the square ji variable, the means of the size 10 samples have a biased distribution to the right. For samples of larger size (n = 100) the approximation to the normal distribution is narrower.

Limit these could not satisfy the level of accuracy with a relatively small sample. If the suppositions on the population are incorrect, then the sampling distribution can be quite inaccurate. In addition, it can be very difficult to deduce mathematically the sampling distribution of the statistical of interest. All these questions can be latent since with only one sample of particular size it seems to be impossible establishing which the sampling distribution is. The idea of the bootstrap focus on this situation: only one data set is available by hand of a certain size and questions if it is possible determining the sampling distribution of its data without using the Theorem of Central Limit. Bootstrapping is a general approach of statistical inference based on these ideas of creating the sampling distribution for a statistics through re-sampling of data close at hand.

The formalization of the bootstrap method is owed to Bradley Efron (Efron 1979 and Efron and Tibshirani 1993) who took ideas from various precursor statistical procedures: the random sampling of finite populations, the estimation of variances from various samplings, stratified semi-samplings and inference methods of intensive calculation, specially Monte Carlo and the jackknife. Efron describes in the above mentioned paper the difficult decision for choosing the name of the method, motivated by the audacity of Tukey (1958) for naming in a special way his methods (jackknife, stem and leaf diagram, box and mustache graphic). Tukey proposed the jackknife by analogy with those big pocket razors that have a great many different tools that are "pulled" so as the user will be capable of solving many small tasks without using some better tool. Statistically speaking, the jackknife is a general approach to prove hypotheses and calculate confidence intervals. It was originally introduced by Quenouille (1949, 1956) as a method of bias reduction, when there were no best methods to be used. This can happen when it is difficult the estimation of statistics, due to the fact that the sampling distributions cannot be exactly deduced or its bias is not known and thus, it is difficult to create confidence intervals. In the case of the bootstrap, the term invented by Efron he refers to someone that "pulls" himself upwards with the bands of its boots. With this strange expression, Efron wanted to reflect the use of the only sample available for giving rise to many others. In this paper the bootstrap will be reviewed as a statistical method of intensive calculation for data analysis. Diverse algorithms will be shown associated with the estimation by bootstrap intervals and its application will be presented with the problem relative to mean estimation. Finally, mention will be median of the scopes and limitations of this method.

Materials and Methods

Procedure of the Bootstrap. It is assumed that a sample is taken $M = \{y_1, y_2, \dots, y_n\}$ of a population

Teorema del Límite Central, éstos no podrían satisfacer el nivel de exactitud con una muestra relativamente pequeña. Si los supuestos acerca de la población son incorrectos, entonces la distribución de muestreo puede ser bastante inexacta. Además, puede ser muy difícil deducir matemáticamente la distribución de muestreo del estadístico de interés. Todas estas interrogantes pueden estar latentes porque ¡con una sola muestra de un tamaño particular parece ser imposible establecer cuál es la distribución de muestreo! La idea del bootstrap gira en torno a esta situación: solamente tenemos a la mano un conjunto de datos de un cierto tamaño y nos preguntamos si es posible determinar la distribución de muestreo de los datos mismos, sin tener que hacer uso del Teorema del Límite Central. El Bootstrapping es un enfoque general de inferencia estadística basado en estas ideas de construir la distribución de muestreo para una estadística a través del remuestreo de los datos a la mano.

La formalización del método bootstrap se debe a Bradley Efron (Efron 1979 y Efron y Tibshirani 1993), tomando ideas de varios procedimientos estadísticos precursores: el muestreo aleatorio de poblaciones finitas, la estimación de varianzas a partir de varios muestreos, semimuestreos estratificados y métodos de inferencia de cómputo intensivo, en especial Monte Carlo y el jackknife. Efron describe en el artículo referido la difícil decisión para elegir el nombre del método, motivado por el atrevimiento de John Tukey (1958) para nombrar de manera peculiar a sus métodos (jackknife, diagrama tallo y hoja, gráfico de caja y bigotes, etc). Tukey propuso el jackknife por analogía con aquellas navajas grandes de bolsillo que tienen multitud de herramientas que se "jalan" para que el usuario sea capaz de resolver muchas tareas pequeñas sin tener que usar alguna herramienta mejor. Estadísticamente hablando, el jackknife es un enfoque general para probar hipótesis y calcular intervalos de confianza (introducido originalmente por Quenouille (1949, 1956) como un método de reducción del sesgo) donde no hubieran mejores métodos para usar. Esto puede ocurrir cuando es difícil la estimación de estadísticos debido a que las distribuciones de muestreo no pueden deducirse exactamente, ni se conocen su sesgo y, por tanto es difícil construir intervalos de confianza. En el caso del bootstrap, el término inventado por Efron hace alusión a alguien que se "jala" a sí mismo hacia arriba con los cintillos de sus botas. Con esta rara expresión Efron quiso reflejar el uso de la única muestra disponible para dar lugar a muchas otras. En este trabajo revisaremos el bootstrap como método estadístico de cómputo intensivo para el análisis de datos. Se mostrarán los diversos algoritmos asociados con la estimación por intervalos bootstrap y se ilustrará su aplicación con el problema relativo a la estimación de la mediana. Finalmente, se hará mención de los alcances y limitaciones de este método

Materiales y Métodos

Procedimiento del Bootstrap. Supóngase que tomamos una muestra $M = \{y_1, y_2, \dots, y_n\}$ de una población

$P = \{y_1 + y_2 \dots y_n\}$, that N is much larger than n , and that M is a simple random sample or an independent random sample of the P population. It is assumed that the elements of the population are scaled (although it can be considered as multivariate data) and that interests some statistics $T = t(M)$ as an estimation of the corresponding population parameter ($\theta = t(P)$). The decision of choosing the bootstrap method (specifically, the non-parametric bootstrap) is due to the fact that the exact T distribution cannot be deduced and neither is it possible using the alternative of asymptotic results because there is no accurate fulfillment in those cases in which there is only a relatively small sample.

For performing the non-parametric bootstrap, no suppositions are made on the population form, but a random sample is selected of the n size of the M sample (as the sample were from an estimation of the population, P), with replacement sampling (for avoiding that only the original M sample stays). The M sample play the role of P population, from which repeated samples are taken, the bootstrap samples. If it is called the first bootstrap sample $M_j^* = \{y_{1j}, y_{2j}, \dots, y_{nj}\}$, then each selected element from M , y_{il}^* from M , will be part of the bootstrap sample, with $1/n$ probability, that is, copying the selection of the M sample of the P population. Repeating this procedure many times for selecting many bootstrap samples, for example N times, it will be obtained, in general, the j -th bootstrap sample $M_j^* = \{y_{1j}, y_{2j}, \dots, y_{nj}\}$. For finding the bootstrap estimation of the T statistics, the following steps are taken:

1. For each M_j^* bootstrap sample, calculate the statistical values $T_j^*, j=1 \dots n$ (If the T_j^* distribution is created around the original T estimation, then this distribution is similar to the sampling T distribution surrounding θ).

2. Estimate the expected value of the T sampling distribution through the estimation of the expected value of the bootstrap estimations, using the mean of the calculated statistics in the bootstrap samples,

$$\hat{E}(T) = \hat{E}^*(T^*) = \bar{T}^* = \frac{\sum_{j=1}^N T_j^*}{N}$$

There is no guarantee that the $\hat{E}(T)$ estimator would be unbiased, so part of the bootstrap procedure is estimating the bias for $\hat{E}(T) = T^*$. On the other hand, the $T, B = T - \theta$ bias can be estimated as $B^* = T^* - T$

3. If you wish to have an estimation of the standard error of the sampling T distribution, calculate the estimation of the standard error of the bootstrap estimations using the standard deviation of the

$$E\hat{E}(T) = E\hat{E}^*(T^*) = \sqrt{\hat{V}^*(T^*)} = \sqrt{\frac{\sum_{j=1}^N (T_j^* - \bar{T}^*)^2}{N-1}}$$

$P = \{y_1 + y_2 \dots y_n\}$, que N es mucho más grande que n , y que es una muestra aleatoria simple o una muestra aleatoria independiente de la población P . Asumamos también que los elementos de la población son escalares (aunque también podrían considerarse datos multivariados) y que estamos interesados en algún estadístico $T = t(M)$ como una estimación del correspondiente parámetro poblacional $\theta = t(P)$. La decisión de elegir el método bootstrap (específicamente, el bootstrap no-paramétrico) obedece a que no podemos deducir la distribución exacta de T y que tampoco es posible usar la alternativa de resultados asintóticos porque no se cumplen con exactitud en aquellos casos en que solamente contamos con una muestra relativamente pequeña.

Para efectuar el bootstrap no-paramétrico no hacemos suposiciones sobre la forma de la población, sino que seleccionamos una muestra aleatoria de tamaño n de la muestra M (como si la muestra se tratara de una estimación de la población, P), muestreando con reemplazo (para evitar quedarnos con solamente la muestra original M). La muestra M juega el papel de la población P de la cual se toman muestras repetidas, las muestras bootstrap. Si llamamos a la primera muestra bootstrap $M_j^* = \{y_{1j}, y_{2j}, \dots, y_{nj}\}$ entonces cada elemento seleccionado de M , y_{il}^* de M , pasará a formar parte de la muestra bootstrap, con probabilidad $1/n$, es decir, imitando la selección de la muestra M de la población P . Repitiendo este procedimiento muchas veces para seleccionar muchas muestras bootstrap, digamos N veces, obtendremos en general, la j -ésima muestra bootstrap $M_j^* = \{y_{1j}, y_{2j}, \dots, y_{nj}\}$. Para hallar la estimación bootstrap del estadístico T llevamos a cabo los siguientes pasos:

1. Para cada muestra bootstrap M_j^* , calcule los valores del estadístico $T_j^*, j=1 \dots n$, (Si construimos la distribución de los T_j^* en torno a la estimación original T entonces esta distribución es análoga a la distribución de muestreo de T alrededor de θ).

2. Estime el valor esperado de la distribución de muestreo de T , a través de la estimación del valor esperado de las estimaciones bootstrap, usando la media de las estadísticas calculadas en las muestras bootstrap,

$$\hat{E}(T) = \hat{E}^*(T^*) = \bar{T}^* = \frac{\sum_{j=1}^N T_j^*}{N}$$

No hay garantía de que el estimador $\hat{E}(T)$ sea insesgado, de manera que parte del procedimiento bootstrap es estimar el sesgo para $\hat{E}(T) = T^*$. Por su parte, el sesgo de $T, B = T - \theta$, se puede estimar como: $B^* = T^* - T$

3. Si se desea tener una estimación del error estándar de la distribución de muestreo de T , calcúlese la estimación del error estándar de las estimaciones bootstrap, usando la desviación estándar de las estadísticas calculadas en las muestras bootstrap,

$$E\hat{E}(T) = E\hat{E}^*(T^*) = \sqrt{\hat{V}^*(T^*)} = \sqrt{\frac{\sum_{j=1}^N (T_j^* - \bar{T}^*)^2}{N-1}}$$

statistics calculated in the bootstrap samples,

Similarly as it happens with $\hat{E}(T)$ there is no guarantee that the estimator of the standard error of T, calculated through the bootstrap samples is accurate. This step could be omitted as in the cases of the estimation of confidence intervals of bootstrap by the percentile methods described below.

If instead of selecting randomly bootstrap samples for obtaining the $\hat{E}^*(T^*)$ and $\hat{EE}^*(T^*)$ estimators, all possible bootstrap samples of n size are numbered for generating all the elements of the $B(n)=\{M_j^*M_j^*\}$ set, it is a bootstrap sample of n size so that $E^*(T^*)$ and $EE^*(T^*)$ can be accurately calculated, this from the computational point of view would be prohibited. The number of possible bootstrap samples (B (n) cardinalship) is very large unless in case n is small. It can be demonstrated that such number is:

$$\text{Card } \mathbf{B}(n) = \binom{2n-1}{n}$$

For example, $\text{Card } \mathbf{B}(15) = 7.8 \times 10^7$; $\text{Card } \mathbf{B}(20) = 6.9 \times 10^{10}$.

In the bootstrap inference, combined to the mistake committed using a particular M sample for representing the population, a second mistake is made by not listing completely all bootstrap samples. This latter mistake can be controlled making a sufficiently large number of bootstrap repetitions.

At first the bootstrap method can be seen as an analogy in which the estimation method is based at times. If F is the distribution of the population that generated a random sample of size n, then the estimator T and its sampling distribution G (T) can be considered as F functions. Efron suggested that it is possible to substitute F by a consistent estimation \hat{E} . This estimated distribution \hat{E} is a function of empirical distribution assigning the sample probability, $1/n$, to each observation in the random sample. Finally what Efron made was to estimate the sampling distribution G(T) with the bootstrap distribution of the statistical values $G^*(\hat{E})$ by Monte Carlo simulation.

Estimation by bootstrap confidence intervals. From its origin, one of the main research tasks in bootstrapping has been the development of methods for calculating valid confidence limits for population parameters (figure 2). These methods are of very diverse kinds and are mainly differentiated by the assumptions made on the estimators and simulation algorithms for obtaining bootstrap samples. Manly (2007) describes some of these methods. The most popular are:

• Standard bootstrap (normal method): It suppose that the bootstrap estimator has an approximately normal distribution and the bootstrap re-sampling gives a good approximation of the statistical standard error of interest.

- First percentile method (percentile method

Análogamente a como sucede con $\hat{E}(T)$ no hay garantía de que el estimador del error estándar de T, calculado a través de las muestras bootstrap, sea exacto. Este paso se pudiera omitir, como en los casos de la estimación por intervalos de confianza bootstrap por los métodos del percentil que se describen abajo.

Uno se preguntará si, en lugar de seleccionar aleatoriamente muestras bootstrap para obtener los estimadores $\hat{E}^*(T^*)$ y $\hat{EE}^*(T^*)$, enumeramos todas las muestras bootstrap posibles de tamaño n, para generar todos los elementos del conjunto $B(n)=\{M_j^*M_j^*\}$ es una muestra bootstrap de tamaño n, de manera que podamos calcular $E^*(T^*)$ y $EE^*(T^*)$ exactamente. Esto sería prohibitivo desde el punto de vista computacional. El número de muestras bootstrap posibles (la cardinalidad de B(n)) es muy grande, a menos que n sea pequeño. Se puede demostrar que tal número es:

$$\text{Card } \mathbf{B}(n) = \binom{2n-1}{n}$$

Por ejemplo, $\text{Card } \mathbf{B}(15) = 7.8 \times 10^7$; $\text{Card } \mathbf{B}(20) = 6.9 \times 10^{10}$.

En la inferencia bootstrap, aunado al error cometido por usar una muestra particular M para representar a la población, se incurre en un segundo error al no enumerar, completamente todas las muestras bootstrap. Este último error se puede controlar haciendo que el número de repeticiones bootstrap sea suficientemente grande

El método bootstrap se puede ver como una analogía al principio detrás del cual se basa el método de estimación por momentos. Si F es la distribución de la población que generó una muestra aleatoria de tamaño n, entonces el estimador T y su distribución de muestreo G(T) pueden considerarse ambas funciones de F. Efron sugirió que es posible reemplazar F por una estimación consistente \hat{E} . Esta distribución estimada \hat{E} es una función de distribución empírica que asigna la misma probabilidad, $1/n$, a cada observación en la muestra aleatoria. Finalmente, lo que hizo Efron fue estimar la distribución de muestreo G(T) con la distribución bootstrap de los valores del estadístico $G^*(\hat{E})$ por simulación Monte Carlo.

Estimación por intervalos de confianza bootstrap. Desde su origen, una de las principales tareas de investigación en bootstrapping ha sido el desarrollo de métodos para calcular límites de confianza válidos para parámetros poblacionales (figura 2). Estos métodos son de muy diversa índole, y se diferencian principalmente por las suposiciones que se hacen sobre los estimadores y los algoritmos de simulación para obtener muestras bootstrap. Manly (2007) describe varios de estos métodos, siendo los más populares:

• Bootstrap estándar (o Método normal): Supone que el estimador bootstrap tiene una distribución aproximadamente normal y el remuestreo bootstrap da una buena aproximación del error estándar del estadístico de interés

• Primer método del percentil (Método del percentil

Use of data and bootstrap distribution for inferring a sampling distribution

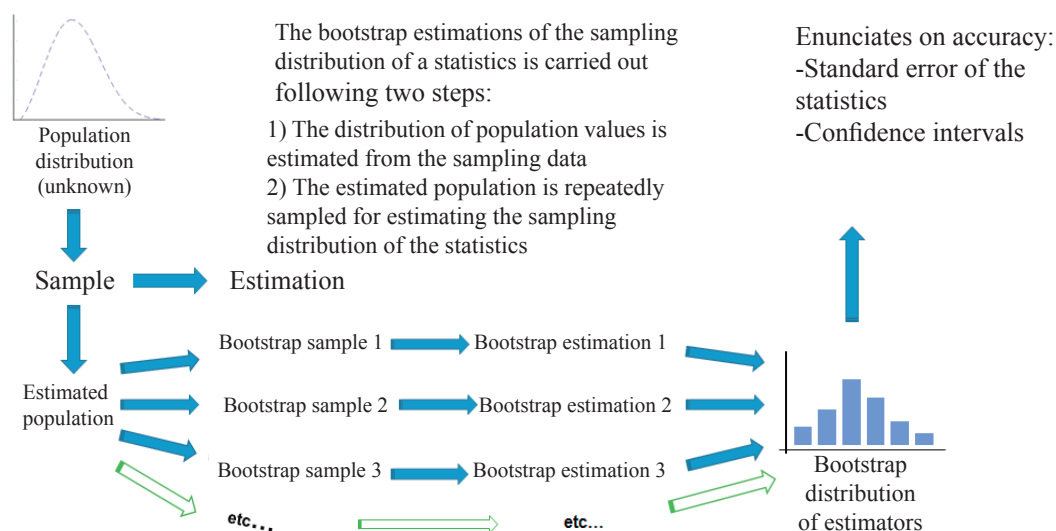


Figure 2. Procedure for the estimation of bootstrap confidence intervals (adapted from Quinn and Keough 2002)

of Efron (1979). The limits of $100(1 - \alpha)\%$ are determined with values of the bootstrap distribution of the statistics of interest occupied by the percentiles $100(\alpha/2)\%$ and $100(1 - \alpha/2)\%$. In this procedure it is supposed that there is a monotonous transformation of estimator values distributed normally.

- Second percentile method (Hall's method). It is based on generating the distribution of the difference between the bootstrap estimation and the estimation of the parameter calculated with the original sample.

- Percentile with corrected bias. It corrects any bias arising from applying the first percentile method, making the median of the estimator distribution be equal to the mean. The algorithm of the percentile method with corrected bias uses a more complex algorithm than the first percentile since in the confidence limits takes into account the proportion of times that the bootstrap estimation is higher than the estimation of the parameter using an original sample.

- Percentile with correction by accelerated bias. Proposed by Efron and Tibshirani (1986), it assumes that there is also a monotonous value transformation of the estimator normally distributed, but the mean and standard error of this distribution are linear functions of the transformation itself as could happen when the standard error of the distribution varies with the mean.

- Bootstrap-t. It uses a t pivotal statistic calculated for each bootstrap sample. The calculations in this algorithm are more intensive, requiring calculating the parameter estimators, as of its standard error, for each bootstrap sample (this standard error could be calculated by jackknife)

Bootstrap confidence limits for the mean. In some statistical texts (e.g. Sokal and Rohlf 2012) is presented the formula: $EE_{\text{sampling median}} = 1.253 \sigma / \sqrt{n}$, where α is the

de Efron (1979)): Los límites del $100(1 - \alpha)\%$ se determinan con los valores de la distribución bootstrap del estadístico de interés que ocupan los percentiles $100(\alpha/2)\%$ y $100(1 - \alpha/2)\%$. En este procedimiento se supone que existe una transformación monótona de valores del estimador que se distribuye normalmente.

• Segundo método del percentil (Método de Hall). Se basa en generar la distribución de la diferencia entre la estimación bootstrap y la estimación del parámetro calculada con la muestra original.

• Percentil con sesgo corregido. Corrige cualquier sesgo que surgiera al aplicar el primer método del percentil, haciendo que la mediana de la distribución del estimador sea igual a la media; el algoritmo del método del percentil con sesgo corregido usa un algoritmo más complejo que el del primer percentil porque en los límites de confianza toman en cuenta la proporción de veces que la estimación bootstrap es mayor que la estimación del parámetro usando muestra original.

• Percentil con corrección por sesgo acelerado. Propuesto por Efron y Tibshirani (1986); asume que existe también una transformación monótona de valores del estimador que se distribuye normalmente, pero la media y el error estándar de esta distribución son funciones lineales de la transformación misma, como podría suceder cuando el error estándar de la distribución varía con la media.

• Bootstrap-t. Usa un estadístico pivotal t que se calcula para cada muestra bootstrap. Los cálculos en este algoritmo son más intensivos pues requiere calcular tanto estimadores del parámetro como de su error estándar para cada muestra bootstrap (este error estándar se puede calcular por jackknife)

Límites de confianza bootstrap para la mediana. En algunos textos de estadística (e.g. Sokal y Rohlf 2012)

population standard deviation of a continuous random variable, Y . If α is not known, then its substitution with the sampling standard deviation would only be valid for large samples and normal distributions. But it could occur that the distribution of variables of the random variable Y does not follow a normal one, making invalid the application of the estimator of the median standard error, and thus, the calculation of the confidence limits for this parameter. In the case of Y variables, positively biased, that follow a distribution log normal, an alternative method can be used involving taking $Y' = \log(Y)$ as variable approximately normal for calculating the conventional confidence interval based on the t of Student for Y' , mean. The median parametric confidence interval will be obtained by applying the transformation backwards the interval for this mean. Given the restrictions for applying these parametric methods of median estimation, the use of non-parametric methods is generally chosen. The most common are:

1. Quantiles of binomial distribution with percentile $p = 0.5$. These quantiles are the same used for the median sign test (Hollander and Wolfe 2011). Given a confidence level $(1 - \alpha) \times 100\%$, the binomial probabilities supply inferior (x') and superior (x) critical values, corresponding to half the significance level, $\alpha/2$. These critical values correspond to $R_{inf} = x' + 1$ and $R_{sup} = n - x' = x$ ranges of the arranged sample so as the values of the variable occupying those places, $y_{R_{inf}}$ and $y_{R_{sup}}$ will constitute the confidence limits of $(1 - \alpha) \times 100\%$ for the mean. This method is adequate for small samples, that is, $n \leq 20$ (Helsel & Hirsch 2002).

2. Approximation of the method of binomial quantiles to the normal distribution. It is preferably applied for large samples ($n > 20$) and it is based on critical values of the standard normal $z_{\alpha/2}$. The confidence limits for the median $y_{R_{inf}}$ and $y_{R_{sup}}$ will be those values of the arranged sample occupying places offered by the following expressions, conveniently rounded off to the closer whole numbers:

$$R_{inf} = (n - z_{\alpha/2} \sqrt{n})/2 \text{ y } R_{sup} = (n + z_{\alpha/2} \sqrt{n})/2 + 1$$

3. Method based on the range test with Wilcoxon sign. It is based on the equivalence of the statistics T^+ of this test with the average number of Walsh positive. Walsh averages for a sample $\{y_1, y_2, \dots, y_n\}$ constitute a set of $n(n+1)/2$, numbers $W_{ij} = (y_i + y_j)/2, i < j$. When Walsh averages are arranged from lesser to greater, $W_{(1)}, W_{(2)}, \dots, W_{(n(n+1)/2)}$, the confidence limits can be calculated following the same procedure of the quantile method of the binomial distribution. The mean of the Walsh averages can be used as estimator (from Hodges-Lehmann) of the distribution median from where the sample was taken under the supposition that such distribution was symmetric. If the distribution is not symmetric, then the estimator will be denominated the pseudo-median.

4. Non-parametric bootstrap. This method is

se presenta la fórmula:

$$EE_{\text{Mediana muestral}} = 1.253\sigma/\sqrt{n},$$

donde α es la desviación estándar poblacional de una variable aleatoria continua, Y . Si α no se conoce, entonces su sustitución con la desviación estándar muestral solamente sería válida para muestras grandes y distribuciones normales. Pero pudiera ocurrir que la distribución de valores de la variable aleatoria Y no siguiera una normal, haciendo inválida la aplicación del estimador del error estándar de la mediana y , por ende, el cálculo de los límites de confianza para este parámetro. Para el caso de variables Y positivamente sesgadas que siguen una distribución lognormal se puede usar un método alternativo que involucra tomar $Y' = \log(Y)$ como variable aproximadamente normal para calcular el intervalo de confianza convencional basado en la t de Student para la media de Y' ; el intervalo de confianza paramétrico de la mediana se obtendrá aplicando la transformación hacia atrás del intervalo para esa media. Dadas las restricciones para aplicar estos métodos paramétricos de estimación de la mediana, generalmente se opta por el uso de métodos no paramétricos. Los más comunes son:

1. Cuantiles de la distribución binomial con percentil $p=0.5$. Estos cuantiles son los mismos que se usan para la prueba del signo de la mediana (Hollander y Wolfe 2011). Dado un nivel de confianza $(1-\alpha) \times 100\%$, las probabilidades binomiales proporcionan valores críticos inferior (x') y superior (x) correspondientes a la mitad del nivel de significación, $\alpha/2$. Estos valores críticos corresponden a rangos $R_{inf} = x' + 1$ y $R_{sup} = n - x' = x$ de la muestra ordenada, de modo que los valores de la variable que ocupan esos lugares, $y_{R_{inf}}$ y $y_{R_{sup}}$ conformarán los límites de confianza del $(1-\alpha) \times 100\%$ para la mediana. Este método es adecuado para muestras pequeñas, digamos $n \leq 20$ (Helsel & Hirsch 2002).

2. Aproximación del método de cuantiles binomiales a la distribución normal. Se aplica preferentemente para muestras grandes ($n > 20$) y se basa en valores críticos de la normal estándar $z_{\alpha/2}$. Los límites de confianza para la mediana $y_{R_{inf}}$ y $y_{R_{sup}}$ serán aquellos valores de la muestra ordenada que ocupan los lugares dados por las siguientes expresiones, convenientemente redondeadas a los enteros más próximos:

$$R_{inf} = (n - z_{\alpha/2} \sqrt{n})/2 \text{ y } R_{sup} = (n + z_{\alpha/2} \sqrt{n})/2 + 1$$

3. Método basado en la prueba de rangos con signo de Wilcoxon. Se basa en la equivalencia del estadístico T^+ de esta prueba con el número de promedios de Walsh positivos. Los promedios de Walsh para una muestra $\{y_1, y_2, \dots, y_n\}$ conforman un conjunto de $n(n+1)/2$ números $W_{ij} = (y_i + y_j)/2, i < j$. Cuando los promedios de Walsh se ordenan de menor a mayor, $W_{(1)}, W_{(2)}, \dots, W_{(n(n+1)/2)}$, los límites de confianza se pueden calcular siguiendo el mismo procedimiento del método de los cuantiles de la distribución binomial. La mediana de los promedios de Walsh puede usarse como estimador (de Hodges-Lehmann) de la mediana de la distribución de donde se tomó la muestra, bajo el supuesto de que

described with the following example: Figure 3 shows the values of DDT concentration, measured in 12 specimens of the American perch, sampled in the Tennessee River (Sincich 1993). In view of the positive bias observed in the sample of DDT concentrations, it is reasonable to use the median as summary measurement of the value distribution center of the variable $Y = \text{"DDT concentration"}$ and presenting a confidence interval for the median for inference purposes.

In total, nine methods were applied. First, the backwards transformation of the confidence interval was utilized for the median of $\log(Y)$, under the supposition that Y is distributed log-normally. For

tal distribución sea simétrica. Si la distribución no es simétrica, entonces al estimador se le denominará la pseudomediana.

4. Bootstrap no-paramétrico. Este método se describe con el siguiente ejemplo: La figura 3 muestra el diagrama de caja de los valores de concentración de DDT medidos en 12 ejemplares de la perca americana, muestreados en el Río Tennessee (Sincich 1993). Debido al sesgo positivo observado en la muestra de concentraciones de DDT, tiene sentido usar a la mediana como medida resumen del centro de la distribución de valores de la variable $Y = \text{"concentración de DDT"}$ y presentar un intervalo de confianza para la mediana para fines de inferencia.

DDT concentrations in American perchs (n = 12)

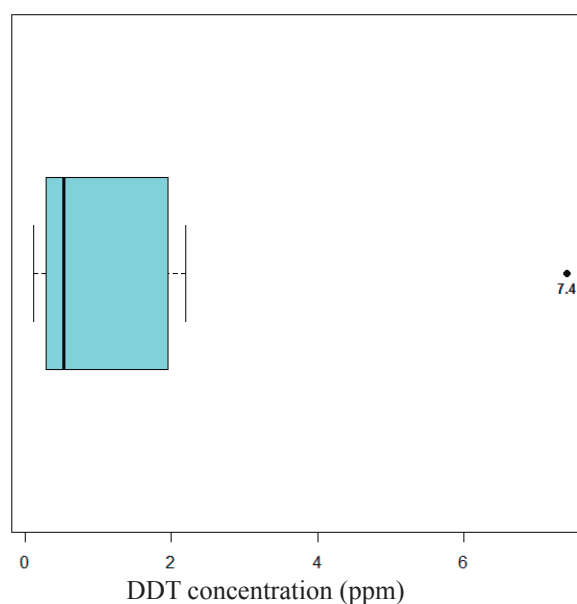


Figure 3. Box plot. DDT concentrations in American perchs. An extreme datum (7.4 ppm) is observed.

avoiding the possibility of assuming any distribution, for the variable Y it was also estimated the confidence interval for the median based on non-parametric methods. Methods 1-3 were applied which were described in the previous section and also the bootstrap (4) was employed, according to the procedure illustrated in figure 2 for inferring the sampling distribution of the sampling median based on the examined sample. For this, the bootstrap distribution (empirical) of the median estimations was generated by simulation, in which each median was estimated from the random selection with substitution of the DDT values for obtaining later the median non-parametric bootstrap confidence limits. The estimation methods used were five: Standard or Normal, First Percentile Method, Second Percentile Method (also called "Ordinary"), Percentile of Correction by Accelerated Bias and Bootstrap-t. The calculations were made with functions of the R program (R Core Team 2015) and the "simpleboot" package (Peng 2008).

Se aplicaron, en total, nueve métodos en primera instancia se aplicó la transformación hacia atrás del intervalo de confianza para la media de $\log(Y)$, bajo el supuesto de que Y se distribuye log-normalmente. Para evitar tener que asumir distribución alguna para la variable Y se estimó también el intervalo de confianza para la mediana basado en métodos no paramétricos. Se aplicaron los métodos 1-3 descritos en la sección anterior y también se aplicó bootstrap (4), siguiendo el procedimiento ilustrado en la figura 2 para inferir la distribución de muestreo de la mediana muestral basado en la muestra observada. Para ello se generó por simulación la distribución bootstrap (empírica) de las estimaciones de la mediana, en donde cada mediana se estimó a partir de la selección aleatoria con reemplazo de los valores de DDT observados, para obtener posteriormente los límites de confianza bootstrap no-paramétricos de la mediana. Los métodos de estimación usados fueron cinco: Estándar o Normal, Primer Método del Percentil, Segundo Método del Percentil (llamado también "Ordinario"), Percentil de corrección por sesgo acelerado y

Results and Discussion

Results from the estimations by intervals for the median concentration of DDT with the use of the nine methods described in the section “Confidence limits for the median”, are set out in table 1. It can be noticed that the normal bootstrap method, the second percentile method and the bootstrap-t produce confidence intervals whose lower limits are negative. However, the parametric space of the median is no-negative since the DDT concentration values are also no-negative. Therefore, such lower limits must be changed to 0. Under this criterion, the smallest confidence interval is generated by the second percentile method. A graphic summary of the estimated confidence limits with this method is shown in figure 4.

Bootstrap-t. Los cálculos se efectuaron con funciones del programa R (R Core Team 2015) y el paquete “simpleboot” (Peng 2008).

Resultados y Discusión

Los resultados de las estimaciones por intervalos para la concentración mediana de DDT, usando los nueve métodos descritos en la sección “Límites de confianza para la mediana”, se presentan en la tabla 1. Se puede observar que el método bootstrap normal, el segundo método del percentil y el bootstrap-t producen intervalos de confianza cuyos límites inferiores son negativos. Sin embargo, el espacio paramétrico de la mediana es no-negativo porque los valores de concentración de DDT son también no-negativos. Por tanto, tales límites inferiores se deben cambiar a 0. Bajo este criterio, el intervalo de confianza más pequeño lo genera el segundo método del percentil. Un resumen gráfico de los límites de confianza estimados con este método se muestra en la figura 4.

Table 1. Estimation by confidence intervals of 95 % for the median of the DDT concentration in American perchs, from a sample size $n = 12$. The estimation methods are grouped in three categories: 1) the parametric assuming a log-normal distribution for the DDT values; 2) the conventional non-parametric with three methods based in ranges; and 3) Bootstrap with five methods. Stated that the parametric (of the median) space is the set of real no-negative numbers, the superior no-negative limits produced by three of the bootstrap algorithms were changed to 0 (in parentheses). IL = Inferior Limit, SL= Superior Limit.

Median of the sample	0.53	
1) 1) Parametric method	IL	SL
log-normal	0.30	1.45
2) Non-parametric methods based on ranges	IL	SL
Quantile binomial	0.22	2.00
Normal approximation to the binomial	0.34	1.90
Wilcoxon of ranges with sign (pseudo-median = 1.0575)	0.33	2.10
3) Bootstrap	IL	SL
Normal method	-0.51 (0)	1.24
First percentile method	0.28	1.95
Second percentile method	-0.89 (0)	0.78
Percentile with correction by accelerated bias	0.26	1.90
Bootstrap-t	-0.59 (0)	1.59

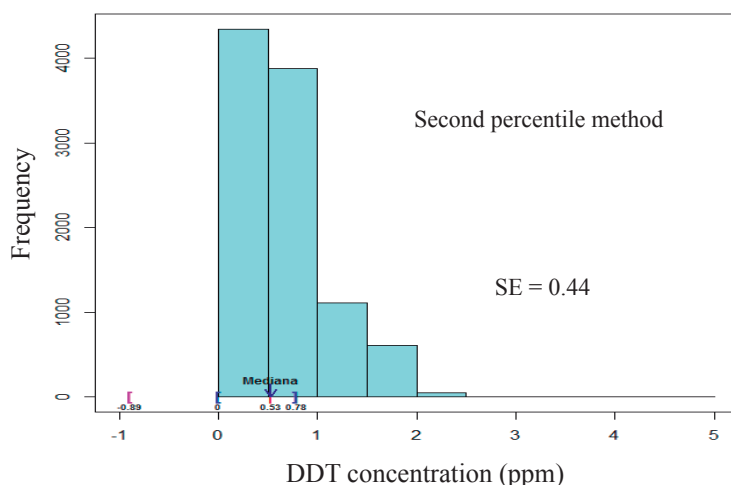


Figure 4. Distribution of bootstrap median frequencies for 10,000 bootstrap samples based on a sample of DDT values of $n = 12$ American perchs and bootstrap confidence interval of 95 % for the median, estimated with the second percentile method. The estimation with negative value (-0.89) of the inferior confidence limit was substituted by 0 which is the minimum possible value of the median parametric space.

Which is the best method (parametric or non-parametric) for estimating confidence intervals for the median? From the set of data analyzed it can be discerned that the functioning of the different estimation methods by intervals for the median is very variable. Surprisingly, the parametric interval exceeds in its accuracy the non-parametric methods. In addition, the percentile methods and the percentile with correction by accelerated bias estimate bootstrap intervals for the median similar to those calculated with conventional non-parametric methods. It also attracts attention the bootstrap intervals producing negative limits, in spite that the parametric space is the set of real no-negative numbers. This brings about the need of modifying the result for obtaining sensible results. What has been observed with this small example of median estimation is recurrent in statistics: there is no clear “winner” as best estimator of some parameter especially when the sample is relatively small. This also applies, in particular, to the non-parametric bootstrap methods. As Manly (2007) indicates it is not possible anticipating in general which of the bootstrap methods allows obtaining better estimations of a parameter. The bright side regarding the use of the bootstrap is that it increases the options for obtaining sensible estimations by parameter intervals of interest in practically any area of applied sciences.

Bootstrap has been used for estimating standard errors and creating confidence intervals of parameters of only one sample, but the procedure can be extended for the estimation of parameters of two or more samples and for bootstrap hypothesis tests. For example, for median comparison of two samples a test statistics (based on the medians of each sample and the overall median) can be selected so as its value is comparable to a bootstrap distribution for which the null hypothesis is true. An alternative approach implies adjusting the sampling values through the residues (the difference between each datum and the sample median to which it belongs). A description of these two approaches of bootstrap tests for two samples can be seen in Manly (2007). In this same reference there is an extensive list of bootstrap application methods in biological sciences, especially in Ecology, Genetics, Evolution, Community Ecology and Environmental Studies.

In the field of agricultural sciences bootstrap is less used but not for that is less important. A recent example is the work of Cubayano *et al.* (2015) who assessed the accuracy of prediction models of reproductive values in beef cattle. These authors used bootstrap for comparing the prediction reliabilities of reproductive values of some complex traits. Another recent paper is that of Narinç *et al.* (2015) comparing two median estimations of three parameters characterizing the color of egg yolks: one was the

¿Cuál es el mejor método (paramétrico o no-paramétrico) para estimar intervalos de confianza para la mediana? Del conjunto de datos analizados se puede observar que el desempeño de los diferentes métodos de estimación por intervalos para la mediana es muy variable. Sorprendentemente el intervalo paramétrico supera en su precisión a los métodos no paramétricos. Además, los métodos del percentil y del percentil con corrección por sesgo acelerado estiman intervalos bootstrap para la mediana, similares a aquellos calculados con métodos no paramétricos convencionales. También llaman la atención los intervalos bootstrap que produjeron límites negativos, a pesar de que el espacio paramétrico es el conjunto de números reales no-negativos. Esto ocasiona que se tenga que modificar el resultado para obtener resultados sensatos. Lo que se ha observado con este pequeño ejemplo de estimación de la mediana es recurrente en Estadística: no hay un claro “ganador” como mejor estimador de algún parámetro, especialmente cuando la muestra es relativamente pequeña. Esto aplica también, en particular, a los métodos del bootstrap no-paramétrico. Como señala Manly (2007) no es posible anticipar en general cuál de los métodos bootstrap permite obtener mejores estimaciones de un parámetro. El lado positivo con relación al uso del bootstrap es que amplía las opciones para obtener estimaciones por intervalos sensatas de parámetros de interés en prácticamente cualquier área de las ciencias aplicadas.

El bootstrap ha sido usado para estimar errores estándar y construir intervalos de confianza de parámetros para una sola muestra, pero se puede ampliar el procedimiento para la estimación de parámetros de dos o más muestras y para pruebas de hipótesis bootstrap. Por ejemplo, para la comparación de las medias de dos muestras se puede elegir un estadístico de prueba (basado en las medias de cada muestra y la media global) tal que su valor se compare con una distribución bootstrap para la cual la hipótesis nula es verdadera. Un enfoque alternativo implica ajustar los valores muestrales a través de los residuos (la diferencia entre cada dato y la media de la muestra a la que pertenece). Una descripción de estos dos enfoques de pruebas bootstrap para dos muestras se puede ver en Manly (2007). En esta misma referencia se ofrecen una extensa lista de aplicaciones de los métodos bootstrap en las ciencias biológicas, especialmente en Ecología, Genética, Evolución, Ecología de Comunidades y Estudios Ambientales.

En el ámbito de las ciencias agropecuarias el uso del bootstrap es menor, pero no por ello es menos importante. Un ejemplo reciente es el trabajo de Cubayano *et al.* (2015) quienes evaluaron la exactitud de modelos de predicción de valores reproductivos en ganado vacuno. Estos autores usaron bootstrap para comparar las confiabilidades de predicciones de los valores reproductivos de varios rasgos complejos. Otro trabajo reciente es el de Narinç *et al.* (2015), en el cual compararon dos estimaciones

conventional median estimation and the other the bootstrap estimation. This type of study, in which the conventional parametric methods and the bootstrap are applied and compared, is frequent in animal science. Hence, in the determination of reference intervals of physiological parameters in animals, the bootstrap confidence intervals seem to have better performance under conditions where the conventional parametric methods are not applicable. Examples of these studies of reference intervals are those carried out by Bennett *et al.* (2006) with physiological parameters in greyhounds and by Cooper *et al.* (2014) with biochemical and hematological parameters in pigs. In other estimation problems, as the determination of the sample size in studies searching for the optimization of the number of samples, the bootstrap method has shown to be a good alternative (see Bravo-Iglesias 2010 and Bravo-Iglesias *et al.* 2013).

The strategy of realizing statistical bootstrap tests for solving problems that have to do with the analysis of variance of one or various factors, the multiple regressions or the survival has been used in numerous agricultural and animal sciences investigations. In the majority of the cases, the bootstrap is utilized for proving the fit goodness of models or its selection (Casellas *et al.* 2006, Sahinler and Karakok 2008 and Tarrés *et al.* 2011, Faridi *et al.* 2014 and Rodríguez *et al.* 2013). Additionally, it has been employed for the control of error rates Type I in multiple tests (Meuwissen and Goddard 2004 on false discovery rates in the comparison of two treatments affecting the expressions of a considerable amount of genes).

Since the publication of the classical papers of Efron (1979), the use of bootstrap has been explosive since many persons have found in the method complete response to difficult questions. The most popular bootstrap method is the one having the simplest theoretical bases: Efron's percentile. The users are more reluctant to use the other methods due to the sophistication of the supporting theory. For example, Manly and Navarro (2009) applied bootstrap-t for estimating the confidence interval for the difference of two medians when the variances are very different and distributions are very biased and these authors found very imprecise results. Therefore, the bootstrap must be utilized careful in situations in which it has not been thoroughly tested. Theory guarantees that the bootstrap will function well in certain situations with large samples, but with small samples it is not possible anticipating if it will operate better or worst (j) than the parametric and conventional non-parametric methods. Prior suggesting the particular utilization of bootstrap for small samples, it is recommended testing the procedure with simulated samples. Undoubtedly the bootstrapping applications for confidence intervals and significance tests will continue developing in the future

de las medias de tres parámetros que caracterizan el color de yemas de huevo: una de ellas fue la estimación convencional de la media y , la otra, su estimación bootstrap. Este tipo de estudios en donde se aplican y comparan los métodos paramétrico convencional y el bootstrap es recurrente dentro de la ciencia animal. Así, en la determinación de intervalos de referencia de parámetros fisiológicos en animales los intervalos de confianza bootstrap parecen tener un mejor desempeño en situaciones en donde los métodos paramétricos convencionales no son aplicables. Ejemplos de estos estudios de intervalos de referencia son los realizados por Bennett *et al.* (2006) con parámetros fisiológicos en galgos y Cooper *et al.* (2014) con parámetros bioquímicos y hematológicos en cerdos. En otros problemas de estimación, tales como la determinación del tamaño de muestra en estudios que buscan optimizar el número de muestras, el método bootstrap ha mostrado ser una buena alternativa (ver, por ejemplo, los estudios de Bravo-Iglesias 2010 y Bravo-Iglesias *et al.* 2013).

La estrategia de realizar pruebas estadísticas bootstrap para resolver problemas que tienen que ver con el análisis de varianza de uno o varios factores, de regresión múltiple, o de supervivencia ha sido utilizada en numerosas investigaciones dentro de la ciencias agropecuarias y ciencia animal. En la mayoría de los casos el bootstrap se usa para probar la bondad de ajuste de modelos o la selección de éstos (Casellas *et al.* 2006; Sahinler y Karakok 2008; Tarrés *et al.* 2011; Faridi *et al.* 2014; Rodríguez *et al.* 2013). También se ha usado para el control de las tasas de error Tipo I en pruebas múltiples (ver el trabajo de Meuwissen y Goddard (2004) sobre tasas de descubrimiento falsas en la comparación de dos tratamientos que afectan las expresiones de una cantidad considerable de genes).

Desde la publicación del artículo clásico de Efron (1979), el uso del bootstrap ha sido explosivo porque muchas personas han hallado en el método la respuesta completa a muchas preguntas difíciles. El método bootstrap más popular es el que posee las bases teóricas más sencillas: el percentil de Efron. Los usuarios son más reacios a usar los otros métodos por la sofisticación de la teoría que los sustenta. Además existen ejemplos en los que bootstrap simplemente no funciona bien. Por ejemplo, Manly y Navarro (2009) aplicaron bootstrap-t para estimar el intervalo de confianza para la diferencia de dos medias cuando las varianzas son muy distintas y las distribuciones son muy sesgadas, observando resultados muy imprecisos. Por tanto, el bootstrap debe usarse con cuidado en situaciones en donde no ha sido probado rigurosamente. La teoría garantiza que el bootstrap funcionará bien en ciertas situaciones con muestras grandes, pero con muestras pequeñas no es posible anticipar si funcionará mejor o peor (j) que los métodos paramétricos y no paramétricos convencionales. Antes de sugerir el uso particular del bootstrap para muestras pequeñas se recomienda probar el procedimiento con muestras simuladas. Sin duda que las aplicaciones del

and only those methods with clear properties and with computer support for its operation in standard statistical packages will prevail.

bootstrapping para intervalos de confianza y pruebas de significación seguirán desarrollándose en el futuro y prevalecerán solamente aquellos métodos con propiedades claras, y con soporte computacional para su uso en paquetes estadísticos estándar.

References

- Bennett, S., Abraham, L., Anderson, G., Holloway, S. & Parry, B. 2006. "Reference limits for urinary fractional excretion of electrolytes in adult non-racing Greyhound dogs". *Australian Veterinary Journal*, 84 (11): 393–397, ISSN: 1751-0813, DOI: 10.1111/j.1751-0813.2006.00057.x.
- Bravo, I. J. A. 2010. Aplicación del método Bootstrap para la estimación de parámetros poblacionales en Parcelas Permanentes de Muestreo y en la modelación Matemática en plantaciones de *Pinus cubensis* Griseb. Ph.D. Thesis, Universidad Pinar del Río, Cuba.
- Bravo, J., Torres V., Rodríguez, L., Montalvo, J., Toirac, W., Fuentes, V. & Rodríguez, P. 2011. "Determinación del tamaño de muestra en parcelas permanentes de muestreo mediante la aplicación del método bootstrap". *Revista Forestal Baracoa*, 32 (1): 57–65, ISSN: 0138-6441.
- Casellas, J., Torres, J., Piedrafita, J. & Varona, L. 2006. "Parametric bootstrap for testing model fitting in the proportional hazards framework: An application to the survival analysis of Bruna dels Pirineus beef calves". *Journal of Animal Science*, 84 (10): 2609–2616, ISSN: 0021-8812, 1525-3163, DOI: 10.2527/jas.2005-729.
- Cooper, C. A., Moraes, L. E., Murray, J. D. & Owens, S. D. 2014. "Hematologic and biochemical reference intervals for specific pathogen free 6-week-old Hampshire-Yorkshire crossbred pigs". *Journal of Animal Science and Biotechnology*, 5, p. 5, ISSN: 2049-1891, DOI: 10.1186/2049-1891-5-5.
- Cuyabano, B. C. D., Su, G., Rosa, G. J. M., Lund, M. S. & Gianola, D. 2015. "Bootstrap study of genome-enabled prediction reliabilities using haplotype blocks across Nordic Red cattle breeds". *Journal of Dairy Science*, 98 (10): 7351–7363, ISSN: 0022-0302, DOI: 10.3168/jds.2015-9360.
- Efron, B. 1979. "Bootstrap Methods: Another Look at the Jackknife". *The Annals of Statistics*, 7 (1): 1–26, ISSN: 0090-5364, 2168-8966, DOI: 10.1214/aos/1176344552, MR: MR515681Zbl: 0406.62024.
- Efron, B. & Tibshirani, R. J. 1994. *An Introduction to the Bootstrap*. CRC Press, 456 p., ISBN: 978-0-412-04231-7.
- Faridi, A., Golian, A., Mousavi, A. H. & France J. 2013. "Bootstrapped neural network models for analyzing the responses of broiler chicks to dietary protein and branched chain amino acids". *Canadian Journal of Animal Science*, 94 (1): 79–85, ISSN: 0008-3984, DOI: 10.4141/cjas2013-078.
- Helsel, D. R. & Hirsch, 2002. *Statistical methods in Water Resources Techniques of Water Resources Investigations*. Book 4, chapter A3. U.S. Geological Survey. Reston, USA.
- Hollander, M. & Wolfe, D. A. 1999. *Non parametric Statistical Methods*. 2nd Ed. Wiley Interscience, New York, 816 p.
- Manly, B. F. J. 2007. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. 3rd ed., Boca Raton, FL: Chapman and Hall/CRC, 480 p., ISBN: 978-1-58488-541-2.
- Manly, B. F. J. & Navarro, J. 2009. "Bootstrap Tests and Confidence Intervals for Non-Normal Data". In: Ronja F. (ed.), *Neue Methoden der Biometrie: Beiträge des 55. Biometrischen Kolloquiums an der Leibniz-Universität Hannover 2009*, Hannover, Germany: Institut für Biostatistik, p. 76.
- Meuwissen, T. H. E. & Goddard, M. E. 2004. "Bootstrapping of gene-expression data improves and controls the false discovery rate of differentially expressed genes". *Genetics, selection, evolution: GSE*, 36 (2): 191–205, ISSN: 0999-193X, DOI: 10.1051/gse:2003058, PMID: 15040898PMCID: PMC2697185.
- Nariñ, D., Aygün, A., Küçükönder, H., Aksoy, T. & Gürcan, E. K. 2015. "An Application of Bootstrap Technique in Animal Science: Egg Yolk Color Sample". *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 21 (5): 631–637, ISSN: 1300-6045, 1309-2251.
- Peng, R. D. 2008. Simpleboot: Simple Bootstrap Routines. R package. version 1.1-3, Available: <https://cran.itam.mx/bin/windows/contrib/3.2/simpleboot_1.1-3.zip>.
- Quenouille, M. H. 1949. "Approximate tests of correlation in time-series 3". *Mathematical Proceedings of the Cambridge Philosophical Society*, 45 (03): 483–484, ISSN: 1469-8064, DOI: 10.1017/S0305004100025123.
- Quenouille, M. H. 1956. "Notes on Bias in Estimation". *Biometrika*, 43 (3/4): 353–360, ISSN: 0006-3444, DOI: 10.2307/2332914.
- Quinn, G. P. & Keough, M. J. 2002. *Experimental Design and Data Analysis for Biologists*. Cambridge, New York: Cambridge University Press, 560 p., ISBN: 978-0-521-00976-8.
- R Core Team. 2015. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, Available: <<https://www.R-project.org/>>.
- Rodríguez, L., Larduet, R., Martínez, R. O., Torres, V., Herrera, M., Medina, Y. & Noda, A. C. 2013. "Modelación de la dinámica de acumulación de biomasa en *Pennisetum purpureum* vs. king grass en el occidente de Cuba". *Revista Cubana de Ciencia Agrícola*, 47 (2): 119–124, ISSN: 2079-3480.
- Sahinler, S. & Karakok, S. G. 2008. "Bootstrap and Jackknife Parameter Estimation of the Models Fitting to Lactation Milk Yield (2x305) on Calving Age". *Journal of Applied Animal Research*, 34 (1): 39–44, ISSN: 0971-2119, DOI: 10.1080/09712119.2008.9706937.
- Sincich, T. 1993. *Statistics by Example*. 5th ed., New York: Dellen Pub Co, 1024 p., ISBN: 978-0-02-410981-1.

- Sokol, R. R. & Rohlf, F. J. 2012. *Biometry: The principles and practice of statistics in biological research*. 4th Ed. W. F. Freeman and Co.
- Tarrés, J., Fina, M., Varona, L. & Piedrafitá, J. 2011. “Carcass conformation and fat cover scores in beef cattle: A comparison of threshold linear models vs grouped data models”. *Genetics Selection Evolution*, 43 (1): 1–10, ISSN: 1297-9686, DOI: 10.1186/1297-9686-43-16.
- Tukey, J. W. 1958. “Bias and confidence in not quite large samples”. *The Annals of Mathematical Statistics*, 29 (2): 614–623, ISSN: 0003-4851, DOI: 10.1214/aoms/1177706647.

Received: June 20, 2015