

## Intervalos de confianza bootstrap del índice de biodiversidad de Shannon

### Bootstrap confidence intervals for Shannon biodiversity index

L. Pla<sup>1</sup> y S. D. Matteucci<sup>2</sup>

#### Resumen

Un aspecto que ha recibido poca atención en los estudios de biodiversidad, es el de asociar a los índices estimados una medida de confiabilidad, una alternativa es el método de remuestreo (bootstrap). Se busca escoger el mejor método para construir intervalos de confianza (IC) bootstrap con probabilidad conocida (1- $\alpha$ ) para el índice de biodiversidad de Shannon ( $H'$ ). Se usaron muestras de seis formaciones de Beard del bosque seco tropical del estado Falcón, Venezuela. Para cada formación, se generaron  $B=5000$  muestras bootstrap para calcular IC con 95% de confiabilidad por los métodos: estándar, de percentiles simples, de percentiles de desvíos, y de percentiles corregidos. La comparación de estos métodos se basó en la amplitud y en la forma del intervalo. Los sitios seleccionados ( $n=185$ ) presentaron valores de  $H'$  desde 0.557 hasta 4.097 correspondientes a riquezas específicas de 8 a 129. La distribución empírica mostró una ligera asimetría positiva. Al comparar los IC obtenidos para las seis formaciones presentes en Falcón, el método que mejor reproduce el carácter asimétrico de la distribución original es el de los percentiles corregidos con parámetros de forma: 0.458 para los Desiertos, 0.501 para los Espinares, 0.548 para las Selvas Veraneras, 0.652 para los Cardonales, 0.659 para los Bosques de Galería y 0.914 para los Arbustales. El método utilizado puede aplicarse también a otros índices de biodiversidad asociando así una medida de confiabilidad al promedio.

**Palabras clave:** remuestreo, formaciones de Beard, bosque seco tropical, Falcón, Venezuela

---

Recibido el 3-6-2000 ● Aceptado el 12-6-2001

1. Departamento de Ambiente y Tecnología Agrícola - Universidad Nacional Experimental 'Francisco de Miranda'. Apdo. 7434, Coro 4101, Venezuela. E-mail: laura@reacciun.ve

2. Investigador independiente de CONICET, Centro de Estudios Avanzados - Universidad de Buenos Aires, J.E. Uriburu 950, 1 piso, Capital Federal 1114, Argentina. E-mail: smatt@arnet.com.ar

## Abstract

Whatever the index used to quantify biodiversity, An aspect that has received little attention in biodiversity studies is that of associating estimated indexes with confidence levels. One option could be the "bootstrap" method. (which do not require normality assumption and may be used with any sample size.) The aim is to choose the best approach to construct bootstrap confidence intervals (BCI) with known probability (1- $\alpha$ ) for the Shannon biodiversity index ( $H'$ ). Six Falcón State dry tropical forest Beard formations were chosen to compare four bootstrap techniques. For each combination, B=5000 bootstrap samples were generated to calculate an CI estimate of 95% for the following methods: the standard BDC, the simple percentile BCI, the percentile difference BCI and the bias corrected BCI percentile. The comparison was based on interval length and shape. The 185 selected sites had biodiversity indexes from  $H'=0.557$  to  $H'=4.097$  and richness from 8 to 129 species. Empirical distribution showed a slightly positive asymmetry. The percentile bias corrected method stood up in practice to the theoretical expectations, and its results are consistent with various sizes of the original sample. The percentile bias corrected procedure reproduces the asymmetric empirical distribution with shape index of: 0.458 for desert, 0.501 for thorn woodland, 0.548 for seasonal forest, 0.652 for cactus scrub, 0.659 for gallery forest and 0.914 for dry evergreen bushland. This method may be applied to other biodiversity indexes in order to establish average confidence levels in point estimation.

**Key words:** Re-sampling, Beard formations, tropical dry forests, Falcón, Venezuela

## Introducción

La biodiversidad específica es una medida de la cantidad de información de la comunidad biótica que junto con los flujos energéticos y los intercambios de materia regulan, el crecimiento, desarrollo y evolución del espacio ecológico. La cuantificación de la biodiversidad para cualquier nivel de organización, es un instrumento heurístico en la gestión ambiental.

De los numerosos índices utilizados para expresar la biodiversidad específica, los más aceptados son aquellos que combinan los dos componentes principales del concepto: el número de especies y la abundancia relativa de cada

una. Los aspectos relacionados con el muestreo (tamaño, forma y número de muestras adecuadas), el cálculo y la interpretación de los diferentes índices, han sido presentados por (9). Entre los índices basados en la abundancia proporcional de las especies se encuentra el de Shannon; pero cualquiera sea el índice a usar, un aspecto que ha recibido poca atención en los estudios de biodiversidad, es el de asociar a esta estimación una medida de confiabilidad. La construcción de intervalos de confianza basados en el estadístico t-Student, suponen un ajuste razonable a la

distribución Normal o la obtención de una muestra suficientemente grande. Ninguna de estas condiciones exigidas por el método, se cumplen en los estudios comparativos de la biodiversidad específica. Una alternativa consiste en utilizar el método bootstrap de remuestreo que no requiere el supuesto de normalidad y acepta muestras de cualquier tamaño para la estimación de intervalos. Esta técnica permite estimar la variabilidad de la muestra a través de la toma de sucesivas muestras con reemplazo, de la muestra original. El bootstrap fue introducido hace 21 años (5) como una técnica computacional intensiva para calcular el error estándar de una estimación. Ha sido revisado recientemente (3) y Manly propone aplicaciones biológicas (10).

El índice de biodiversidad específica de Shannon, se basa en suponer que la heterogeneidad depende del número de especies presentes y de su abundancia relativa en una comunidad. Conceptualmente, es una medida del grado de incertidumbre asociada a la selección aleatoria de un individuo en la comunidad. Esto es, si una comunidad de S especies es muy homogénea, por ejemplo porque existe una especie claramente dominante y las restantes S-1 especies apenas presentes, el grado de incertidumbre será más bajo que si todas las S especies fueran igualmente abundantes. O sea, al tomar aleatoriamente un individuo en el primer caso tendríamos un grado de certeza más alto (menos incertidumbre) que en el segundo; porque, mientras en el primero, la probabilidad de que pertenezca a la especie dominante será cercana a uno,

mayor que para cualquier otra especie; en el segundo, será la misma para cualquier especie. Si se desea cuantificar este grado de incertidumbre, es posible apoyarse en la teoría de la información suponiendo una población infinita.

Sobre esta base el índice de Shannon se define como (13):

$$H' = - \sum_{i=1}^S p_i \ln p_i$$

siendo la proporción de la i-ésima especie en una comunidad de S especies. La interpretación del antilogaritmo de H', o sea de  $e^{H'}$  cuantifica el número de especies, igualmente abundantes, suficiente para producir el mismo grado de incertidumbre, o sea el mismo valor de H'.

El valor de H' se estima utilizando datos de la abundancia a campo de cada especie como:

$$\hat{H}' = - \sum_{i=1}^S \hat{p}_i \ln \hat{p}_i$$

siendo  $\hat{p}_i = n_i / N$ , o sea un cociente entre la abundancia de la i-ésima especie en la muestra y la abundancia total. Este método produce un estimador sesgado (8) con esperanza matemática

$E(\hat{H}') = - \sum_{i=1}^S p_i \ln p_i - \frac{S-1}{2N} + \frac{1 - \sum_{i=1}^S p_i^{-1}}{12N^2} + \frac{\sum_{i=1}^S (p_i^{-1} - p_i^{-2})}{12N^3} + \dots$  aunque en la práctica, y especialmente si no se dejó de incluir alguna especie en la muestra, este sesgo no es importante (12).

Se trata de construir un intervalo de confianza con probabilidad conocida de (1-a) para el parámetro de interés  $\theta$ , para el cual existe un estimador calculado a partir de una muestra aleatoria  $x_1, x_2, \dots, x_n$ , tomada de la población de interés, con una distribución de probabilidad F desconocida.

El método aplicado para construir los IC puede extenderse a otros índices de biodiversidad y de calidad del ambiente. Además, este método evita cometer el error frecuente que se produce al asociar medidas de confiabilidad a la estimación de un parámetro, cuya distribución es desconocida (y definitivamente no es

normal) y para el que no se dispone de muestras grandes. El objetivo de este trabajo es seleccionar la mejor técnica para construir intervalos bootstraps del índice de biodiversidad de Shannon ( $H'$ ), que permitan la comparación estadística de la diversidad específica de diferentes comunidades.

## Materiales y métodos

Se describen brevemente los cuatro procedimientos para construir intervalos de confianza bootstrap (ICB) que se comparan en este trabajo.

**El método ICB estándar:** es el que más se asemeja a la manera de construir IC paramétricos se supone que el estimador bootstrap  $\hat{\theta}^*$  se distribuye en forma normal con media  $\theta$  y desviación estándar  $F$  y por eso existe una probabilidad  $(1 - \alpha)$  tal que

$$(q + z_{\alpha/2}s < \hat{q}^* < q + z_{(1-\alpha/2)}s)$$

aplique para una muestra aleatoria de la cuál provenga  $\hat{\theta}^*$  (5). Obsérvese que para el límite inferior se ha indicado un valor de  $z$  correspondiente a la distribución normal estandarizada de media cero y varianza unitaria, con una probabilidad  $\alpha/2$ , que corresponde al valor de  $z$  que deja por debajo  $100(\alpha/2)\%$  de la distribución. Esta notación es especialmente útil en casos de remuestreo, ya que coincide con los percentiles de una distribución empírica que reproduzca la teórica. Para el caso de una distribución simétrica respecto a cero, se cumplirá que el negativo del percentil  $100(\alpha/2)\%$

es igual al percentil  $100(1-\alpha/2)\%$ , y por ello es que el IC se encuentra corrientemente expresado con las constantes  $\pm z_{\alpha/2}s$ .

En términos de un intervalo de confianza se expresa como:

$$\Pr [\hat{q}^* + z_{\alpha/2}\hat{s}^* < q < \hat{q}^* + z_{(1-\alpha/2)}\hat{s}^*] = 1 - \alpha$$

usando  $\bar{\theta}^*$  para indicar la media de las estimaciones bootstrap del parámetro y  $\hat{\sigma}^*$  para indicar la desviación estándar de esta estimación. Así,

$$\bar{q}^* = \frac{1}{B} \sum_{j=1}^B \hat{q}_j^* \quad \text{y} \quad \hat{s}^{*2} = \frac{1}{B} \sum_{j=1}^B (\hat{q}_j^* - \bar{q}^*)^2$$

siendo  $B$  el número de muestras bootstrap tomadas.

**El método ICB del percentil simple:** supone la existencia de una transformación monotónica de  $\hat{\theta}$ , sea  $f(\hat{\theta})$  normalmente distribuida con media  $f(\theta)$  y desviación estándar unitaria; aunque esta transformación no sea necesariamente conocida (5).

Si existe  $f(\hat{\theta})$  que tiene distribución  $N[f(\theta), 1]$  entonces podremos calcular los IC partiendo de que se cumplirá

$$(f(\hat{q}) + z_{\alpha/2} < f(q) < f(\hat{q}) + z_{(1-\alpha/2)})$$

con probabilidad  $(1-\alpha)$ , que podemos

ordenar en función de nuestro interés centrándolo en la función del parámetro como

Si se conoce la forma de la transformación  $f(\cdot)$ , sólo hay que invertir la transformación y encontrar los límites de confianza para el parámetro  $\varrho$  con la misma probabilidad. Como no siempre se la conoce, podemos suponer que es generada por remuestreo. Es decir que podemos encontrar una  $f(\hat{\theta}^*)$ , luego generar un gran número de estos valores y construir la distribución empírica a cualquier nivel de exactitud. Con esta distribución es posible encontrar los extremos inferior  $\{f(\hat{\theta}^*) + z_{\alpha/2} = LI\}$  y superior  $\{f(\hat{\theta}^*) + z_{(1-\alpha/2)} = LS\}$ , y como la transformación es monotónica,  $f(\cdot)$  es reversible y pueden encontrarse los límites del intervalo en términos de los  $\hat{\theta}^*$ . Así, los límites para el parámetro  $\varrho$ , no son más que los límites para la estimación bootstrap  $\hat{\theta}^*$  que excede  $100(\alpha/2)\%$  para el inferior y  $100(1-\alpha/2)\%$  para el superior. En estos casos se recomienda no utilizar menos de 1000 muestras bootstraps.

**Método ICB por percentil de los desvíos:** se sustenta en el supuesto de que la distribución generada por remuestreo debe reproducir la distribución de interés (7). Si se define la distribución del error de  $\hat{\theta}$  como  $\theta = \hat{\theta} - \theta$ , este debería ser aproximado por el error de la distribución en el remuestreo, sea  $\hat{\varepsilon}^* = \hat{\theta}^* - \theta$ . Con esta distribución podemos encontrar dos percentiles de interés para los errores  $\varepsilon_L$  y  $\varepsilon_S$  tales que

$$\Pr [ e_L < \hat{q} - q < e_S ] = 1 - \alpha$$

y utilizarlos para construir el IC para el parámetro de interés, como:

$$\Pr [ \hat{q} - e_S < q < \hat{q} - e_L ] = 1 - \alpha$$

Este método se aplica tomando la diferencia para el cálculo de los límites del intervalo respecto al valor calculado en la muestra original o población bootstrap. Al programar el algoritmo debe atenderse a mantener los signos de los percentiles estimados, de lo contrario se obtendrán resultados contradictorios (límite inferior mayor que el superior, o la inversa).

La distribución bootstrap participa de manera distinta en los cálculos del ICB del percentil simple y del ICB del percentil de los desvíos. Es posible demostrar (10) que mientras el primer método equivale a encontrar un IC con límites en  $\hat{\theta}_{\alpha/2}^*$  y  $\hat{\theta}_{(1-\alpha/2)}^*$ , el segundo corresponde a los límites entre  $2\hat{\theta} - \hat{\theta}_{\alpha/2}^*$  y  $2\hat{\theta} - \hat{\theta}_{(1-\alpha/2)}^*$ . Esta segunda forma de expresar el intervalo es la que permite que se manifieste el carácter asimétrico potencial de la distribución de  $\hat{\theta}$ , siendo particularmente útil cuando no es posible suponer simetría.

Una manera de expresar esta propiedad de asimetría del IC es a través de un cociente o índice de forma (4) que se define como

$$\text{Índice de forma} = (LS - \hat{\theta}^*) / (\hat{\theta}^* - LI)$$

cualquiera sea el método con el que se calculen los límites. Dada la simetría que supone el método del percentil simple, el índice de forma será siempre la unidad.

Al usar el método del percentil de desvíos se estima el parámetro utilizando el percentil 50 de distribución de los errores:  $\varepsilon_{B \rightarrow 0.50}$ . La variable pivote del intervalo será  $\hat{\theta} - \hat{\varepsilon}_{0.5}$ , que

permite calcular el índice de forma como el cociente

$$\begin{aligned} \text{Índice de forma} &= (\text{LS} \cdot (\hat{\theta} - \hat{\epsilon}_{0.5})) \\ &/ ((\hat{\theta} - \hat{\epsilon}_{0.5}) - \text{LI}) \\ &\text{que reduce, luego de algo de álgebra a} \\ &= ((\hat{\theta} - \hat{\epsilon}_{0.975}) - (\theta - \hat{\epsilon}_{0.5})) / ((\theta - \hat{\epsilon}_{0.5}) - \\ &(\hat{\theta} - \hat{\epsilon}_{0.975})) \\ &= (\hat{\epsilon}_{0.50} - \hat{\epsilon}_{0.025}) / (\hat{\epsilon}_{0.975} - \hat{\epsilon}_{0.50}) \end{aligned}$$

reproduciendo la posible asimetría en la distribución del parámetro. No debe llamar la atención la inversión de los límites en el cociente, ya que un índice de forma mayor a la unidad en los propios desvíos, que indicaría asimetría hacia la izquierda, corresponde a una asimetría a la derecha en la variable en estudio.

La calidad en la estimación del ICB puede conocerse al comparar la cobertura teórica del intervalo con la obtenida en remuestreos sucesivos. Se ha demostrado que si se utilizan 1000 muestras bootstrap para calcular ICB por percentiles de desvíos con 95% de confianza, la cobertura real estará entre 93.6 y 96.4 con probabilidad 0.95; y si B=10000 la cobertura mejora a 94.6-95.4% (2).

**Método ICB por percentiles**

**corregidos:** en el método de los percentiles simple se supone que la transformación  $f(\hat{\theta})$  existe y se distribuye en forma normal, en cuyo caso la mediana (o el percentil 50) era un 'buen pivote' para calcular el intervalo de confianza. El supuesto de normalidad implica que:

$$\text{Pr} [ f(\hat{\theta}) > f(\zeta) ] = 0.50$$

equivale a:

$$\text{Pr} [ \hat{\theta} > \theta ] = 0.50 ;$$

sin embargo, esto no es necesariamente cierto.

El método de los percentiles

corregidos hace uso de la información proveniente de las muestras bootstrap para representar la propia distribución bootstrap y corrige cualquier sesgo debido a que el verdadero parámetro no coincide con la mediana; de allí su nombre (6). Por ejemplo, si se calculan 1000 muestras bootstrap y en 550 de ellas, el valor estimado del parámetro estuvo por debajo del valor del parámetro calculado en la muestra original (o población bootstrap), la proporción correspondiente al verdadero valor del parámetro será  $550/1000 = 0.55 (p^*)$ , y no  $500/1000 = 0.50$  (mediana), y esta información sobre la posible asimetría del estimador será útil para mejorar la estimación de los ICB.

Sea una transformación monotónica  $f(\hat{\theta})$  normalmente distribuida con media  $f(\zeta) - z^*$  y desviación estándar igual a 1. Si existe  $f(\hat{\theta}) \sim N ( f(\zeta) - z^*, 1 )$ , entonces los ICB son calculables como

$$z_{\alpha/2} < f(\hat{\theta}) - f(\zeta) + z^* < z_{(1-\alpha/2)}$$

siendo como antes,  $z_{(\alpha/2)}$ ,  $z_{(1-\alpha/2)}$  y  $z^*$  los valores de la distribución normal estandarizada con probabilidades  $\alpha/2$ ,  $(1-\alpha/2)$  y  $p^*$ . Este ICB puede reordenarse como

$$z_{\alpha/2} - f(\hat{\theta}) - z^* < - f(\zeta) < z_{(1-\alpha/2)} - f(\hat{\theta}) - z^*$$

$$- z_{\alpha/2} + f(\hat{\theta}) + z^* > f(\zeta) > - z_{(1-\alpha/2)} + f(\hat{\theta}) + z^*$$

$$- z_{(1-\alpha/2)} + f(\hat{\theta}) + z^* < f(\zeta) < - z_{\alpha/2} + f(\hat{\theta}) + z^*$$

$$f(\hat{\theta}) + z^* - z_{(1-\alpha/2)} < f(\zeta) < f(\hat{\theta}) + z^* - z_{\alpha/2}$$

que expresado en términos del parámetro será

$$\hat{\theta} + z^* - z_{(1-\alpha/2)} < \zeta < \hat{\theta} + z^* - z_{\alpha/2}$$

o en una forma que imita al IC paramétrico como

$$\hat{\theta} + z^* + z_{\alpha/2} < \alpha < \hat{\theta} + z^* + z_{(1-\alpha/2)}$$

Cuando  $p^*$  es 0.50 y por lo tanto  $z^* = 0$ , la expresión equivale a la del percentil simple.

La distribución bootstrap expresada en términos de la transformación  $f(\hat{\theta})$  permite encontrar el valor bootstrap estimado de la función del estimador del parámetro que deja la proporción  $p_i$  de observaciones por debajo y el que deja la proporción  $p_s$  de observaciones por debajo. Estas dos proporciones serán los estimadores bootstraps de las probabilidades tales que

$$\Pr [ f(\hat{\theta}_s^*) < f(\hat{\theta}) + z^* + z_{(1-\alpha/2)} ] = p_s \text{ y}$$

$$\Pr [ f(\hat{\theta}_i^*) < f(\hat{\theta}) + z^* + z_{(\alpha/2)} ] = p_i .$$

Como  $f(\hat{\theta})$  sigue la distribución normal, es posible encontrar el límite superior como

$$\Pr [ f(\hat{\theta}_s^*) - f(\hat{\theta}) < z^* - z^* + z^* + z_{(1-\alpha/2)} ] = p_s$$

$$\Pr [ f(\hat{\theta}_s^*) - f(\hat{\theta}) + z^* < 2 z^* + z_{(1-\alpha/2)} ] = p_s$$

$$\Pr [ f(\hat{\theta}_i^*) - (f(\hat{\theta}) - z^*) < 2 z^* + z_{(1-\alpha/2)} ] = p_s$$

$$\Pr [ Z_s < 2 z^* + z_{(1-\alpha/2)} ] = p_s$$

mientras que para el inferior resultará

$$\Pr [ Z_i < 2 z^* + z_{(\alpha/2)} ] = p_i$$

Los límites del ICB corregido se podrán encontrar a partir de los percentiles  $100p_i$  % y  $100p_s$  % de la distribución bootstrap del estimador del parámetro.

La cobertura de intervalos bootstraps realizados con 0.95 de probabilidad nominal y 1000 simulaciones por este método osciló entre (87-97)%, (92-96)%, (87-95)%, (87-98)%, (90.5-96)% y (88.5-95)%, en un estudio de estimación del parámetro de concentración para diferentes diseños de dilución (1).

El número de remuestreos indicados para los diferentes métodos se establece en mil o más, a excepción del método de los intervalos estándar para el cual 100 réplicas son suficientes.

En este trabajo se usaron 185 muestras de seis formaciones de Beard del bosque seco tropical del estado Falcón, Venezuela para calcular el índice de biodiversidad vegetal de Shannon. Esta información, recolectada a campo entre 1978 y 1982, ha sido utilizada para relacionar la biodiversidad con factores ambientales (11). La información sobre biodiversidad disponible para cada formación, la estimación de la biodiversidad promedio, y el intervalo de confianza basado en t-Student se sintetizan en la cuadro 1. Los valores de la biodiversidad específica para el conjunto de  $n=185$  censos osciló entre 0.557 y 4.097, mientras que la riqueza (S) abarcó desde 8 hasta 129 especies.

Para cada formación de Beard se generaron  $B=5000$  muestras bootstrap para calcular intervalos de confianza con 95% de confiabilidad por los métodos: estándar, de percentiles simples, de percentiles de desvíos, y de percentiles corregidos. Para ello se construyeron cuatro subrutinas (disponibles solicitándolas al primer autor) que se utilizaron con el programa Resampling Stats Version 3.14 (Julian L. Simon ©).

La comparación entre los métodos de construcción del intervalo de confianza se realizó en función de la amplitud del intervalo (diferencia entre límite superior y límite inferior) y de la forma del intervalo (cociente entre la amplitud por encima y por debajo del parámetro estimado) como indicador del carácter asimétrico de la distribución.

**Cuadro 1. Intervalos de confianza paramétricos en las formaciones de Beard**

Formación de Beard	Tamaño de la muestra original	$\bar{H}$	Límites del IC	
			inferior	superior
Selva veranera	74	2,7466	2,6322	2,8609
Espinar	61	2,3479	2,2447	2,4410
Cardonal	14	1,9883a	1,7525	2,2241
Desierto	16	1,7905a	1,4943	2,0867
Arbustal	12	1,7552a	1,5117	1,9622
Bosque de Galería	8	1,6074a	1,1379	2,0768

$\bar{H}$ : Índice de Shannon promedio estimado en la muestra original, misma letra indica superposición de los IC con 95%; IC: intervalo de confianza basado en t-Student.

## Resultados y discusión

La distribución empírica de los Índices de Shannon mostró una ligera asimetría positiva; como ejemplo, se presenta el caso de la Selva Veranera. En la figura 1 se ha representado el histograma de los valores observados con clases de 0.182 unidades, sobre-imponiéndole la distribución normal esperada. La prueba de bondad de ajuste permite aceptar el modelo normal como buen descriptor ( $p < 0.15$ ), aunque se ponen de manifiesto discrepancias que tornan la distribución asimétrica.

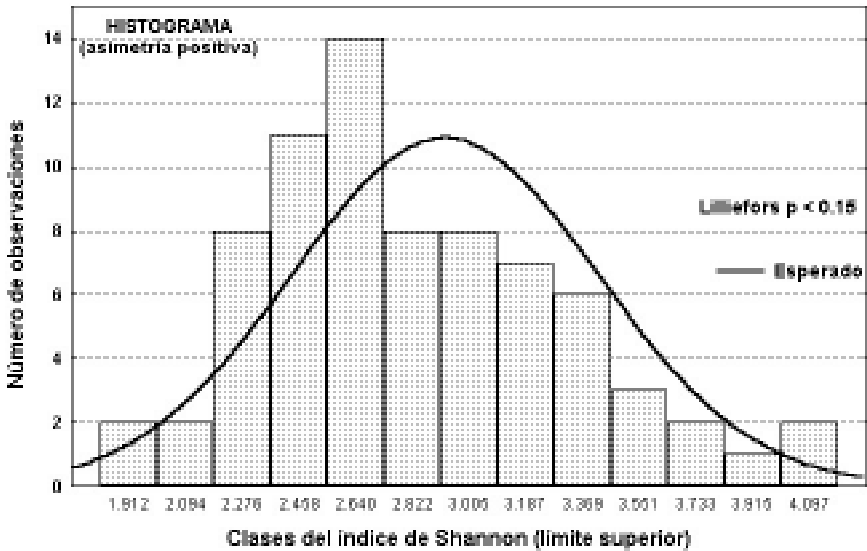
El índice de Shannon promedio en la Selva Veranera observado en la población bootstrap ( $n=74$ ) fue  $H' = 2.7466$ ; la desviación estándar fue  $s_{H'} = 0.49357$ , con máximo y mínimo en  $H'_M = 4.097$  y  $H'_m = 1.730$ ; y un error estándar  $= 0.0574$ . Si suponemos normalidad, el intervalo de confianza de 95% para el índice promedio será () que convenientemente evaluado corresponde a ( $2.6322 < H' < 2.8609$ ). La amplitud de este IC es por lo tanto 0.228.

El logaritmo de la riqueza observada corresponde al índice de Sh-

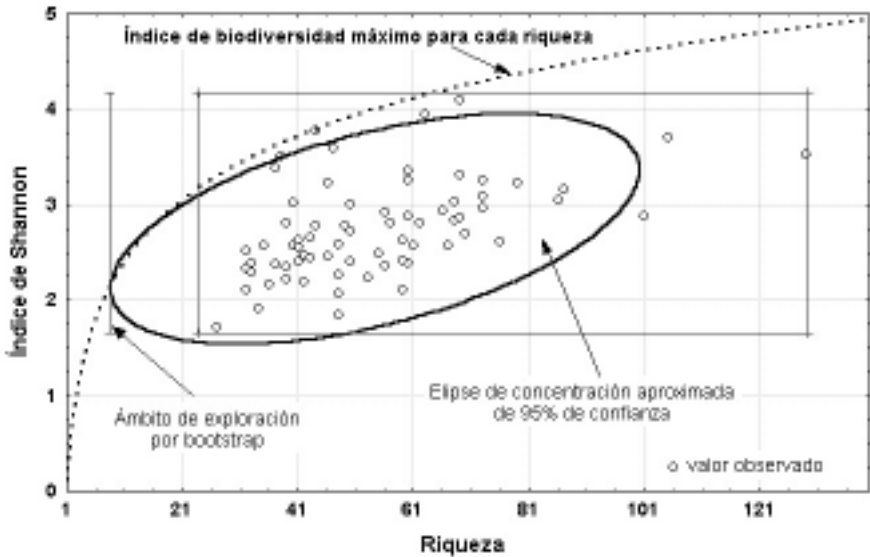
annon esperado en condiciones de máxima equidad (línea de puntos en la figura 2). Así, la figura 2 puede ser usada para interpretar los desvíos horizontales ( $s_{ob} - s_{es}$ : diferencias entre la riqueza observada y la esperada en condiciones de máxima equidad) que indican el número de especies presentes en una comunidad que no ha alcanzado su equidad máxima respecto a las necesarias para que el índice de Shannon observado fuera máximo. Si se interpretan los desvíos verticales ( $H'_{es} - H'_{ob}$ : diferencias de índices de Shannon), se encontrará cuánto de la diversidad alcanzable con equidad máxima no se expresó en la comunidad observada. Se han definido los desvíos de manera que sean siempre positivos para facilitar su interpretación y se ha iniciado el eje de la riqueza en 1 ya que en toda comunidad habrá al menos una especie presente, con índice de Shannon cero: la biodiversidad es nula porque existe una única especie.

Para esta misma formación, la





**Figura 1. Distribución observada y ajuste al modelo Normal en la Selva Veranera de Falcón, Venezuela**



**Figura 2. Población bootstrap de índices de biodiversidad de la Selva Veranera de Falcón, Venezuela**

media bootstrap estándar permitió estimar  $= 2.7478$  con un ICB de 95% correspondiente a  $(2.6347 < H' < 2.8609)$ . La amplitud de 0.2262 resulta muy similar a la estimada en forma paramétrica utilizando la aproximación normal. También el valor promedio fue muy similar y el histograma resultante del remuestreo mostró una forma simétrica.

La media bootstrap, calculada como el percentil 50 de la distribución de muestreo, con el método del percentil simple fue de 2.7435 y el correspondiente intervalo con 95% de confianza resultó  $(2.6366 < H' < 2.8561)$ . El parámetro de forma del intervalo fue 1.0533, indicando una leve asimetría.

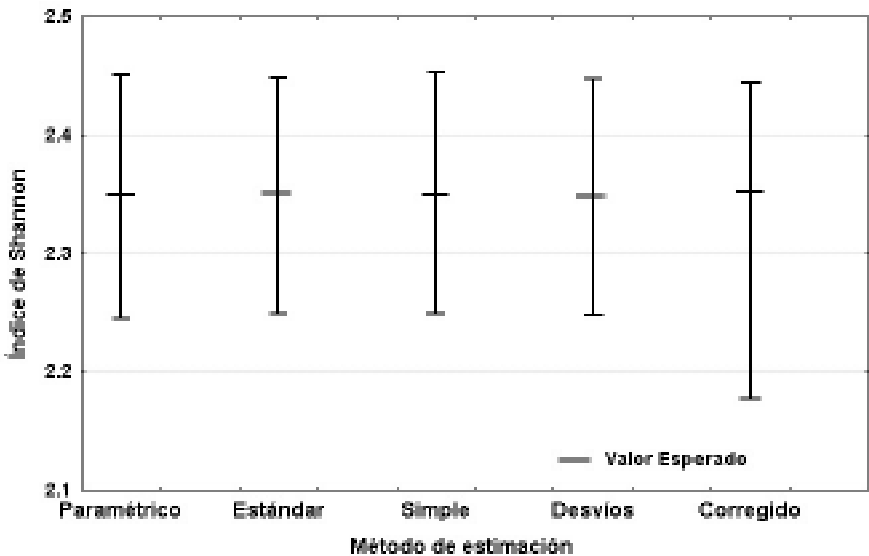
El intervalo calculado por el método del percentil de los desvíos fue  $(2.6323 < H' < 2.8530)$  con estimado en 2.7478. El parámetro de forma fue 0.9108 indicando que el límite supe-

rior se encuentra levemente más cerca del parámetro estimado que el inferior, a diferencia del caso anterior donde la asimetría es a la izquierda.

El resultado obtenido con el método del percentil corregido fue  $(2.5342 < H' < 2.8611)$  con parámetro estimado en el percentil 52.574 correspondiente a un valor de 2.7454. El parámetro de forma (0.5478) indicó que el límite inferior se encuentra más alejado de la media que el superior, señalando la asimetría hacia la derecha de la distribución original.

Los intervalos resultantes de cada método (figura 3) muestran que sólo el que corrige los percentiles estimados reproduce el carácter asimétrico de la distribución original.

Resultados similares a los descriptos hasta aquí se obtuvieron al aplicar estas técnicas y análisis a las restantes formaciones de Beard (cuadro 2).



**Figura 3. Intervalos de confianza bootstrap para la Selva Veranera de Falcón, Venezuela**

**Cuadro 2. Comparación de la amplitud de los intervalos de confianza para las distintas formaciones**

Formación de Beard	Método bootstrap	$\hat{H}''$	Límite del IBC		Forma
			inferior	superior	
Selva veranera	Estándar	2.7478	2.6347	2.8609	-
	Percentil simple	2.7435	2.6366	2.8561	1.0533
	Percentil de desvíos	2.7478	2.6323	2.8530	0.9108
	Percentil corregido	2.7454	2.5342	2.8611	0.5478
Espinar	Estándar	2.3486	2.2488	2.4484	-
	Percentil simple	2.3484	2.2489	2.4532	1.0526
	Percentil de desvíos	2.3476	2.2471	2.4475	0.9932
	Percentil corregido	2.3485	2.1627	2.4415	0.5006
Cardonal	Estándar	1.9865	1.7827	2.1903	-
	Percentil simple	1.9849	1.7843	2.1895	1.0190
	Percentil de desvíos	1.9886	1.7838	2.1919	0.9925
	Percentil corregido	1.9881	1.6691	2.1961	0.6523
Desierto	Estándar	1.7873	1.5234	2.0512	-
	Percentil simple	1.7928	1.5166	2.0373	0.8857
	Percentil de desvíos	1.7871	1.5417	2.0694	1.1501
	Percentil corregido	1.7951	1.2869	2.0282	0.4580
Arbustal	Estándar	1.7530	1.5756	1.9305	-
	Percentil simple	1.7506	1.6004	1.9510	1.3337
	Percentil de desvíos	1.7622	1.5620	1.9090	0.7334
	Percentil corregido	1.7508	1.5206	1.9612	0.9139
Bosque de Galería	Estándar	1.6060	1.2414	1.9706	-
	Percentil simple	1.6074	1.2440	1.9721	1.0038
	Percentil de desvíos	1.6096	1.2404	1.9636	0.9590
	Percentil corregido	1.6040	1.0340	1.9798	0.6592

$\hat{H}''$ : Índice de Shannon estimado por cada método bootstrap  
 ICB: intervalo de confianza bootstrap

## Conclusiones y recomendaciones

Dado que el índice de Shannon es usado ampliamente como indicador de los cambios en la diversidad en función de los cambios de factores ambientales, se requiere una medida aceptable de confiabilidad del valor estimado.

Los intervalos bootstrap son una alternativa cuando se dispone de muestras pequeñas y cuando se desea preservar el carácter asimétrico de la distribución original.

En el caso de las formaciones de Beard del bosque seco tropical de Falcón, las conclusiones que se obtienen por el método paramétrico y el método bootstrap seleccionado difieren a nivel de los límites de los intervalos y si se usa el segundo método no se detectan diferencias

significativas entre espinares y cardonales, cuyos IC paramétricos no se solapan (figura 4).

El método que mejor reproduce el carácter asimétrico de la distribución original es el de los percentiles corregidos. Este método muestra en la práctica las ventajas teóricas que se han postulado y su resultado es consistente con diversos tamaños de la muestra original.

Se recomienda utilizar el método bootstrap de percentiles corregidos para estimar los intervalos de confianza de la media del índice de Shannon. Así, con el mismo esfuerzo experimental, se mejora la robustez de las estimaciones contribuyendo a elevar la calidad de las evaluaciones ambientales.

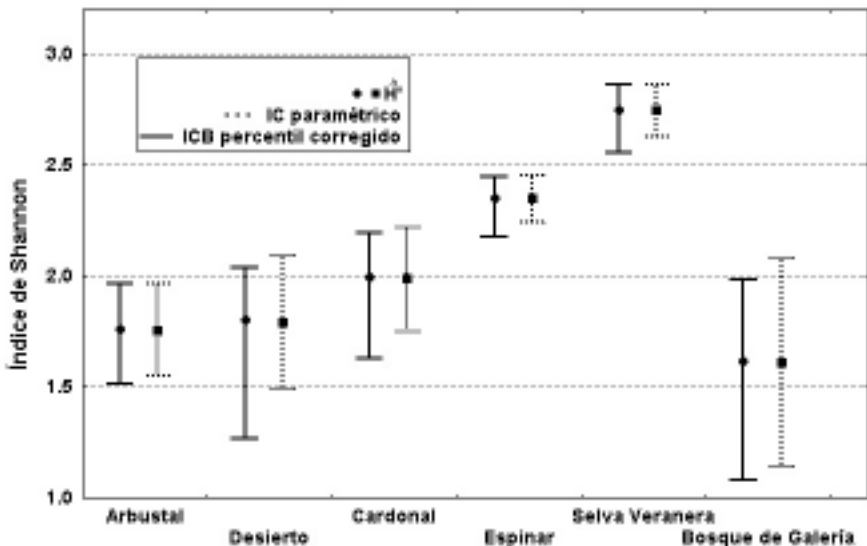


Figura 4. Comparación de intervalos de confianza (95%) en las formaciones de Beard de Falcón, Venezuela

## Literatura citada

1. Basu, R., R. Guerra, y R. Read. 1996. Bootstrap confidence intervals for concentration parameters in dilution assays. *Journal of Agricultural, Biological and Environmental Statistics* 1: 454-466.
2. Buckland, S.T. 1984. Monte Carlo confidence intervals. *Biometrics* 40: 811-817.
3. Davison, A.C. y D.V. Hinkley. 1998. *Bootstrap methods and their application*. Cambridge University Press. New York.
4. DiCiccio, T.J. y B. Efron. 1996. Bootstrap confidence intervals. *Statistical Science* 11: 189-228.
5. Efron, B. 1979. Bootstrap methods: another look at the jackknife. *Annals of statistics* 7: 1-26.
6. Efron, B. 1981. Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics* 9: 139-172.
7. Hall, P. 1992. *The bootstrap and edgeworth expansion*. Springer Verlag, New York.
8. Hutcheson, K. 1970. A test for comparing diversities index based on Shannon formula. *Journal of Theoretical Biology* 29: 151-154.
9. Magurran, A.E. 1988. *Ecological diversity and its measurements*. Chapman & Hall, London.
10. Manly, B.F.J. 1997. *Randomization, bootstrap and Monte Carlo methods in biology*. Second Edition, Chapman & Hall, London.
11. Matteucci, S.D., A. Colma, y L. Pla. 1999. Biodiversidad vegetal en el árido falconiano (Venezuela). *Interciencia* 24: 300-307.
12. Peet, R.K. 1974. The measurement of species diversity. *Annual Review of Ecology and Systematics* 5: 285-307.
13. Shannon, C. y W. Weaver. 1949. *The mathematical theory of communication*. University of Illinois Press, Urbana.