

Facultad de Ciencias

Departamento de Estadística

Carrera Ingeniería Estadística

PROYECTO DE TÍTULO II

| | | |
|---------------------------|---|--|
| Asignatura | : | PROYECTO TÍTULO II (220068) |
| Título | : | Intervalos de confianza Bootstrap para índices de diversidad |
| Realizado por | : | Ana Fica Villalobos Miguel Lagos Barrientos Estudiantes Ingeniería Estadística |
| Profesor Guía | : | Sr. Luis Cid Serrano Departamento de Estadística |
| Profesor Corrector | : | Sra. Nelly Gómez Fuentealba Departamento de Estadística |
| Semestre | : | I Semestre 2014 |

UNIVERSIDAD DEL BÍO BÍO

www.ubiobio.cl

INDICE DE CONTENIDOS

| | |
|--|-----------|
| Introducción..... | 3 |
| Antecedentes Generales..... | 4 |
| Planteamiento del Problema..... | 6 |
| Objetivos..... | 6 |
| Objetivo General..... | 6 |
| Objetivos Específicos..... | 6 |
| Justificación y Alcances de la Investigación..... | 7 |
| Justificación..... | 7 |
| Alcances..... | 8 |
| Marco Conceptual..... | 8 |
| Índices de Diversidad..... | 11 |
| Índice de Shannon..... | 12 |
| Índice de Simpsons..... | 14 |
| Datos y Metodología..... | 15 |
| Estimación por intervalos Bootstrap..... | 16 |
| Fundamentos de la Técnica Bootstrap..... | 18 |
| Estimación de I.C. Bootstrap por Método de Percentiles..... | 19 |
| Estimación de I.C. Bootstrap Corregido por Sesgo y Acelerado..... | 20 |
| Resultados | 22 |
| Conclusiones Generales..... | 35 |
| Referencias..... | 39 |

1. Introducción

Existe una dificultad para definir claramente la diversidad y es que ésta consiste principalmente de dos componentes: la variedad (o riqueza) y la abundancia relativa de especies (u homogeneidad). Adicionalmente, en muchos casos, algunas comunidades que nos podrían interesar estudiar tienen una gran cantidad de individuos y/o no es imposible llegar a todos los miembros por el lugar donde están ubicadas.

Las comunidades tienen estructuras que son diversas, y para ellas tenemos un conjunto de índices que deben ser capaces de representar dichas estructuras comunitarias de distinta forma.

En el proyecto de Título I se resolvió la duda de cómo y cuáles de estos índices nos sirven para un tipo de comunidad específica, qué va a depender para que en una estructura comunitaria usemos un índice u otro; ello dependerá particularmente de las propiedades que tenga dicho índice y que tan bueno sea para representar a dicha comunidad, debido a esto podemos plantear la siguiente pregunta: ¿cómo se mide la capacidad de entrega de información de este índice?

Una respuesta a esta interrogante, se puede encontrar en un intervalo de confianza con un nivel de significación adecuado para cada índice, lo que será analizado en el presente estudio. Estos intervalos no están definidos en la literatura actual, ya que biólogos y ecologistas (quienes son los que más los utilizan) solo entregan estimaciones puntuales a partir de información muestral.

Desde el momento que se trabaja con muestras, ello se transforma en un problema estadístico y la solución se encuentra en la teoría de la estimación de estos intervalos, en nuestro caso lo enfrentaremos usando técnicas de remuestreo (bootstrap).

En el presente estudio nos enfocaremos en analizar los índices de Simpson y Shannon-Wiener. Para ello construiremos dos tipos de Intervalos de Confianza Bootstrap por dos métodos distintos y observaremos sus diferencias. Los intervalos de Confianza seleccionados son: Intervalos de Confianza por el Método del Percentil e Intervalos de Confianza por el método de Sesgo Corregido (BCA).

2. Antecedentes Generales.

La diversidad se define como “la variabilidad entre los organismos vivientes de todas las fuentes, incluyendo, entre otros, los organismos terrestres, marinos y de otros ecosistemas acuáticos, así como los complejos ecológicos de los que forman parte; esto incluye diversidad dentro de las especies, entre especies y de ecosistemas” (UNEP, 1992).

La falta de definición y de parámetros adecuados para su medición hasta principios de los 70's llevó incluso a declarar la falta de validez de este concepto. Actualmente el significado y la importancia de la biodiversidad no están en duda y se han desarrollado una gran cantidad de parámetros para medirla como un indicador del estado de los sistemas ecológicos, con aplicabilidad práctica para fines de conservación, manejo y monitoreo ambiental.

El número o riqueza de especies es la medida más frecuentemente utilizada, por varias razones (Gaston, 1996):

- La riqueza de especies refleja distintos aspectos de la biodiversidad.
- A pesar de que existen muchas aproximaciones para definir el concepto de especie, su significado es ampliamente entendido.
- Al menos para ciertos grupos, las especies son fácilmente detectables y cuantificables.
- Aunque el conocimiento taxonómico no es completo (especialmente para grupos como los hongos, insectos y otros invertebrados en zonas tropicales) existen muchos datos disponibles sobre números de especies.

Uno de los problemas ambientales que han suscitado mayor interés mundial en esta década es la pérdida de biodiversidad como consecuencia de las actividades humanas, ya sea de manera directa (sobreexplotación) o indirecta (alteración del hábitat). Los medios de comunicación han impactado de tal manera que, tanto algunos gobiernos, iniciativas privadas, como la sociedad en general, consideran

prioritario dirigir mayores esfuerzos hacia programas de conservación. Ello debe considerar, prioritariamente, un análisis objetivo de la biodiversidad y su cambio, considerando básicamente una correcta evaluación y monitoreo.

En una reflexión, Wilson (1988) intenta dejar clara la importancia del estudio de la diversidad, él dice: “Sobre los seres vivos con los que compartimos la Tierra desconocemos su número y qué tipo de cosas pueden ofrecernos. Es como una biblioteca con libros sin leer, y ni siquiera hemos acabado el primer capítulo..... estamos perdiendo las especies que hay a nuestro alrededor antes de que podamos pasar a la página siguiente”. Podemos obtener mucho conocimiento al estudiar cómo están compuestas las comunidades de los seres con los cuales co-habítamos. Por ejemplo, algunos diseños de experimentos con pastizales naturales (Cedar Creek) han mostrado la validez de la Hipótesis de Darwin que plantea, “comunidades de vegetales más diversos son más productivos” y además se encontró que se optimiza el uso de nutrientes y de nitrógeno. Los argumentos para su preservación son muchísimos y pasan desde asuntos económicos, evolutivos y hasta morales.

Dentro de las comunidades es posible identificar patrones de variación como el número de especies, número de individuos de una especie, número de individuos de la comunidad y proporción de individuos de una determinada especie en la comunidad. Con ellos es posible construir funciones llamadas índices de diversidad, donde quedan caracterizadas las relaciones de abundancia de especies. Existen muchas formas de calcular estos índices y se han desarrollado gran cantidad de ellos, aproximadamente 140, su aplicación abarca un gran abanico de áreas y sectores, desde poblaciones, individuos, cromosomas, genes, reinos, familias, géneros, bioregiones, paisajes, etc.

3. Planteamiento del Problema

Una vez obtenidos los índices de diversidad, surge la necesidad de generar intervalos de confianza para dichos índices, incorporando de esta forma, una medida de incertidumbre para ellos. Este aspecto es poco analizado en la actualidad existiendo los medios y el conocimiento para realizarlo.

4. Objetivos

4.1 Objetivo General

Construir mediante el método de remuestreo (Bootstrap) dos tipos de intervalos de confianza para dos de los Índices de Diversidad más utilizados en estudios de estructuras de comunidades y comentar sus características.

4.2 Objetivos Específicos

- Obtener mediante Método Bootstrap estimadores para los dos índices de diversidad considerados.
- Estudiar propiedades estadísticas de los índices de diversidad considerados.
- Obtener mediante Método Bootstrap Intervalos de confianza de Percentiles y Sesgo Corregido (BCA) para los índices considerados.

5. Justificación y Alcances de la Investigación.

5.1 Justificación

Para definir claramente la diversidad, es necesario considerar dos componentes: la variedad (o riqueza) y la abundancia relativa de especies (u homogeneidad). Sin embargo, en algunos casos, algunas comunidades contienen una gran cantidad de individuos o es imposible llegar a todos los miembros por el lugar donde están ubicadas constituyendo su evaluación, un problema de muestreo adicional.

Las comunidades tienen estructuras comunitarias distintas, es decir una simetría especial que se refiere a la posición de los individuos unos frente a otros y otra temporal que explica la alternancia de posición de unos y otros, y tenemos un conjunto de índices que deben ser capaces de representar estas estructuras.

Los índices mencionados en la literatura tanto por biólogos o ecologistas (quienes son los que más los utilizan) solo entregan estimaciones puntuales a través de información muestral. Desde el momento que ellos trabajan con muestras, la información no es posible generalizarla a una población y el problema de hacerlo se hace netamente estadístico y la solución es la teoría de la estimación de intervalos y la obtención de estos intervalos se puede realizar en base a técnicas de remuestreo (bootstrap), dado que las propiedades distribucionales de los índices son, en general desconocidas.

El Proyecto de Título II consistirá en realizar estimación de parámetros mediante intervalos de confianza bootstrap por dos métodos, el Método de los Percentiles y el Método de Sesgo corregido (BCA). Para esto utilizaremos el programa estadístico R-Project, con el cual simularemos la población a estudiar que consistirá en una cantidad de especies y el número de individuos por especie. Mediante técnicas de remuestreo calcularemos los índices de diversidad y sus respectivos intervalos de confianza Bootstrap.

5.2 Alcances

El presente Proyecto de Título II tendrá por objetivo obtener intervalos de confianza para los índices de diversidad Shannon-Wiener y Simpson, que aun no se han desarrollado ampliamente. Éste tendrá un alcance a nivel general, para cualquier tipo de comunidad de distintas especies, y proporcionará una metodología para calcular estos intervalos.

6. Marco Conceptual

Los estudios sobre medición de diversidad se han centrado en la búsqueda de parámetros para caracterizarla como una propiedad emergente de las comunidades ecológicas. Sin embargo, las comunidades no están aisladas en un entorno neutro. En cada unidad geográfica, en cada paisaje, se encuentra un número variable de comunidades. Por ello, para comprender los cambios de la biodiversidad con relación a la estructura del paisaje, la separación de los componentes alfa, beta y gamma (Whittaker, 1972) puede ser de gran utilidad, principalmente para medir y monitorear los efectos de las actividades humanas.

Con el propósito de unificar significados de algunos conceptos utilizados en el presente proyecto de título y con el objeto de facilitar la comprensión de este, a continuación se definen los siguientes términos:

Diversidad: la variabilidad entre los organismos vivos de todas las fuentes, incluyendo, entre otros, los organismos terrestres, marinos y de otros ecosistemas acuáticos, así como los complejos ecológicos de los que forman parte; esto incluye diversidad dentro de las especies, entre especies y de ecosistemas.

La **diversidad alfa** es la riqueza de especies de una comunidad particular a la que consideramos homogénea.

La **diversidad beta** es el grado de cambio o reemplazo en la composición de especies entre diferentes comunidades en un paisaje.

La **diversidad gamma** es la riqueza de especies del conjunto de comunidades que integran un paisaje, resultante tanto de las diversidades alfa como de las diversidades beta.

Comunidad: Conjunto de distintas poblaciones de organismos que conviven y se interrelacionan en un lugar determinado.

Especie: Una especie es un grupo de poblaciones capaces de combinarse entre sí pero no con otro grupo similar de poblaciones sobre la base de afinidad y co-dirección en cuanto a especialización ecológica (Shaposhnikov, 1966).

Individuos: Integrantes del grupo de especies estudiadas.

Esta forma de analizar la diversidad resulta muy conveniente en el contexto actual ante la acelerada transformación de los ecosistemas naturales, ya que un simple listado de especies para una región dada no es suficiente. Para monitorear el efecto de los cambios en el ambiente es necesario contar con información de la diversidad biológica en comunidades naturales y modificadas (diversidad alfa) y también de la tasa de cambio en la biodiversidad entre distintas comunidades (diversidad beta), para conocer su contribución al nivel regional (diversidad gamma) y poder diseñar estrategias de conservación y llevar a cabo acciones concretas a escala local.

De acuerdo con (Smith, 2002) a lo largo de las últimas cuatro décadas se han registrado avances importantes en los estudios ecológicos y particularmente de estadística en la ecología, ejemplos de ello son las contribuciones al estudio de las distribuciones espaciales de organismos y los diseños de estudios ecológicos.

La implementación de políticas que incorporan caracterizaciones del estado de una comunidad ecológica, hacen uso de indicadores propuestos varias décadas antes, los *índices de riqueza, equilibrio y diversidad*.

Conviene resaltar la importancia de que la toma de datos se base en un diseño muestral apropiado. Es necesario tener réplicas de cada muestra para poder acompañar el valor de un índice con el de alguna medida de la dispersión de los datos (varianza, desviación estándar o coeficiente de variación), o estimar

el valor mínimo y máximo hipotéticos del índice bajo las condiciones del muestreo (Spellerberg, 1991).

Un aspecto crítico del análisis es asegurarse de que las réplicas estén apropiadamente dispersas (en el espacio o en el tiempo) de acuerdo con la hipótesis que está siendo probada. Esto evita caer en el error señalado por Hurlbert (1984) como pseudoreplicación, que implica la prueba del efecto de algún tratamiento con un término de error inapropiado.

En los análisis de diversidad, esto puede deberse al espacio físico real sobre el cual son tomadas las muestras, o a que las mediciones son inadecuadamente pequeñas, es decir, son restringidas a un espacio menor al inferencial implícito en la hipótesis.

Por lo general se usa la riqueza de especies (R), en su expresión más sencilla, el número de especies distintas de una comunidad, como un sinónimo de diversidad (D). Sin embargo, el concepto de diversidad no sólo contempla una lista de especies distintas, sino también la abundancia de estas especies y su relación con las demás especies. En otras palabras, un índice de diversidad es una función que permite representar la diversidad de especies en una comunidad. La diversidad tiene dos componentes esenciales, *el número de especies* y sus *abundancias relativas*. Una mayor diversidad quedará determinada entonces por un mayor número de especies y una mayor homogeneidad de la distribución de los individuos entre estas especies.

Se deben considerar al menos 3 elementos para describir un índice de diversidad;

- Ningún indicador o combinación de ellos es capaz de describir exhaustivamente las propiedades cuantitativas de la comunidad.
- Los indicadores corresponden a una visión, por lo general instantánea de la comunidad y por tanto, no dan cuenta de los procesos temporales asociados a dicha especie.
- Ambos aspectos anteriores están fuertemente condicionados a las estrategias y diseños muestrales utilizados para extraer la información de interés, en base a la cual se estiman los indicadores.

7. Índices de Diversidad

Los índices de diversidad proveen más información de la composición comunitaria que simplemente la riqueza de especies (número de especies presentes), pues también toman en cuenta la abundancia de las diferentes especies. Si por ejemplo, consideramos dos comunidades de 100 individuos, cada una compuesta de 10 diferentes especies y tal que una de estas comunidades tiene 10 individuos de cada especie y la otra, 9 especies con un individuo y 91 individuos en una décima especie. ¿Cuál de las dos comunidades es más diversa? Aunque poseen el mismo número de especies distintas (su riqueza es similar), la dominancia de una especie por sobre el resto, en la segunda comunidad, la hace menos diversa.

Al tomar en cuenta las abundancias relativas, un índice de diversidad no depende solamente de la riqueza de especies, sino también de la distribución de individuos entre las diferentes especies.

La diversidad es entonces una función de los siguientes parámetros poblacionales s: Número de especies en la muestra.

n_i : Número de individuos de la especie i en la muestra.

N : Número total de individuos en la muestra.

$p_i = \frac{n_i}{n}$: Proporción de individuos de la especie i en la muestra.

Así, podemos definir la función de diversidad D por

$$D = f(s + 1, n, p),$$

La cual debe tener, para ser considerada como un índice de diversidad, las mismas propiedades que la función poblacional.

Entonces, un índice de diversidad es una función

$$D = f(s, n_i, p_i)$$

Que tiene las siguientes propiedades

D es máximo cuando $p_i = 1/s$, para todo i

Si se tienen dos comunidades completamente homogéneas, una con s especies y otra con $s + 1$, entonces D , es una función monótona respecto del número de especies; es decir,

$$f(s + 1, n, p) > f(s, n_i, p_i)$$

Los índices proveen información sobre qué tan rara o común es una especie en una comunidad. La posibilidad de cuantificar la diversidad es una importante herramienta para tratar de entender la estructura de las comunidades naturales.

A continuación presentaremos dos de los índices de diversidad más comúnmente utilizados para describir estructuras comunitarias, y que utilizaremos en este trabajo para desarrollar los métodos de estimación por intervalos, el Método de Percentiles y Sesgo corregido. En general los índices están caracterizados por dos componentes, una la riqueza específica, directamente relacionada con el número de especies y dos, la homogeneidad de la distribución de los individuos entre las especies.

7.1 Índice de Shannon

Es el índice más frecuentemente utilizado en la actualidad, está basado en la teoría de la información, es decir, considera que la información contenida en una comunidad es equivalente a la incertidumbre de la predicción sobre la especie a que pertenecerá un individuo encontrado aleatoriamente. Así, una comunidad con muchas especies en la cual el número de individuos decrece lentamente desde el más, al menos abundante, será considerada de alta información y por consecuencia de una diversidad también alta. El índice tiene la siguiente forma

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

Donde

S : número total de especies en la comunidad.

p_i : Proporción de individuos de la especie i en la comunidad (respecto al total de individuos n_i/N).

n_i : Número de individuos de la especie i .

N : número de todos los individuos de todas las especies.

Si calculamos H_{max} , como la máxima heterogeneidad posible en la comunidad, suponiendo que todas las especies tienen el mismo número de individuos, mediante la expresión:

$$H_{max} = -H'/\ln S$$

La razón proporciona una medida de la homogeneidad de la comunidad. En medida que dicha razón se aproxima a 1, más homogénea será la distribución de los individuos entre las especies.

Valores más altos de H' representan mayor incertidumbre en la identificación de un individuo extraído al azar desde la comunidad. Es claro que esta medida corresponde al grado de incertidumbre acerca de la especie a que pertenece el vecino más cercano a cualquier individuo de la comunidad lo que corresponde por tanto a una medida del grado de su organización, considerando la situación como fruto de una serie de interacciones dentro del propio ecosistema. De esta forma, el índice contempla la cantidad de especies presentes en el área de estudio (*riqueza de especies*), y la cantidad relativa de individuos de cada una de esas especies (*abundancia*).

Este índice es regularmente poco sensible a los cambios en el número de especies raras en la comunidad, esta propiedad de robustez, si bien en determinadas circunstancias pudiera parecer altamente favorable, puede resultar negativa dependiendo de la importancia relativa de cada especie para la comunidad. El uso de este índice supone, en principio, que se trata de una población de tamaño tal que conceptualmente pueda ser considerado como de tamaño infinito (Cid, 2001).

7.2 Índice Simpson

Este índice está fuertemente influido por la importancia de las especies más dominantes (Magurran, 1988). También conocido como el índice de la diversidad de las especies o índice de dominancia y es uno de los parámetros que nos permiten medir la riqueza de organismos. En ecología es también usado para cuantificar la biodiversidad de un hábitat. Considera el número de especies presentes en el hábitat y su abundancia relativa. El índice de Simpson representa la probabilidad de que dos individuos, dentro de un hábitat, seleccionados al azar pertenezcan a especies diferentes.

Para estimarlo, suponemos que se trata de una población de tamaño infinito, compuesta por s especies diferentes, de modo que cada una de ellas tenga $N_1, N_2, N_3, \dots, N_s$ individuos respectivamente

$$D = 1 - \sum_{i=1}^s (p_i)^2$$

Donde:

D/s : es la medida de Simpson para la homogeneidad.

p_i : es la proporción de individuos de la especie i .

El supuesto inicial que la población es infinita, implica que, es prácticamente imposible encuestarla en su totalidad, por tanto, debemos proceder a realizar una estimación del parámetro en base a los resultados de una muestra.

Supongamos entonces, que antes se extrae una muestra de n individuos desde una comunidad con las características señaladas de antemano, y sean $n_1, n_2, n_3, \dots, n_s$ el número de individuos de la muestra que pertenecen a cada una de las especies, de modo que $\sum n_i = n$, un estimador insesgado de D es

$$D = d = \frac{\sum_{i=1}^s n_i(n_i - 1)}{n(n - 1)}$$

8. Datos y Metodología.

Debido a que no disponemos de bases de datos para estudios comunitarios, considerando diferentes condiciones distribucionales – respecto de la abundancia de las especies y de la distribución de los individuos entre ellas – se generaron en este estudio bases de datos simuladas con diferentes distribuciones, tratando de simular distintas condiciones de riqueza específica y de homogeneidad de la distribución de los individuos. Para este efecto, se utilizaron tres distribuciones de probabilidad, las que representan tres distintos tipos de estructuras comunitarias; esto es, los índices considerados son el Índice de Simpson y de Shannon en los tipos de comunidades aquellas que tengan dominancia, es decir algunas especies con mayor número de individuos que otras, las homogéneas, con todas la especies de igual número de individuos y las comunidades heterogéneas.

Además, para estos tres distintos tipos de comunidades consideramos las siguientes distribuciones muestrales: Para las comunidades homogéneas tomamos una distribución uniforme con parámetros 0 y 1, para las comunidades con dominancia le asignamos una distribución exponencial con parámetro $\lambda = \frac{1}{50}$ y para las comunidades heterogéneas trabajamos con la distribución Normal con media 50 y con desviaciones estándar 25, teniendo en cuenta que el valor del coeficiente de variación sea mayor que 0,35. Respecto de los tamaños muestrales, se consideró comunidades desde 10 hasta 100 especies.

En el software estadístico R-projec se creó un programa para calcular intervalos de confianza de sesgo corregido y de percentiles. Se generaron 90 distintas comunidades con número de especies entre 10 y 100, es decir $n_{10} = 10, n_{11} = 11, \dots, n_{100} = 100$ con sus respectivos parámetros, luego se aplicó el programa para el cálculo de los intervalos y se realizó una figura para observar el comportamiento del índice y su intervalo de confianza para cada una de las comunidades y para cada uno de los métodos de intervalos.

8.1 Estimación por intervalos Bootstrap.

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).

El bootstrap es un tipo de técnica de remuestreo de datos que permite resolver problemas relacionados con la estimación de intervalos de confianza o la prueba de significación estadística. Este enfoque puede resultar de interés para los investigadores en diversidad, no solo porque es menos restrictivo que el enfoque estadístico clásico, sino también porque es más general en su formulación y más simple de comprender en lo referente al procedimiento básico que subyace al método. En lugar de fórmulas o modelos matemáticos abstractos, el bootstrap simplemente requiere un computador capaz de simular un proceso de muestreo aleatorio de los datos. Sin embargo, y debido quizás a la escasa difusión de la técnica, los investigadores aún no han incorporado el bootstrap al repertorio habitual de herramientas de análisis de datos

Se busca escoger el mejor método para construir intervalos de confianza (IC) bootstrap con probabilidad conocida, para los índices de diversidad más utilizados.

- **Bootstrap**

El bootstrap es un método de remuestreo no paramétrico propuesto por Efron (1979). La idea central de este método es simple; dada una muestra aleatoria con n observaciones, dicha muestra es tratada como si fuera toda la población de la cual extraeremos m muestras con reemplazo.

Para cada una de las m muestras, se calculará el valor del estimador bootstrap que se utilizará para estimar la variabilidad muestral. Estudios teóricos han demostrado que este enfoque proporciona una buena aproximación de la distribución de los estimadores, lo cual permitirá describir algunas de sus propiedades muestrales, así como el cálculo de intervalos de confianza y la realización de contrastes de hipótesis.

El procedimiento bootstrap es útil para:

- Valorar el sesgo y el error muestral de un estadístico calculado a partir de una muestra.
- Establecer un intervalo de confianza para un parámetro estimado.
- Realizar una prueba de hipótesis respecto a uno o más parámetros poblacionales.

Adicionalmente, el bootstrap tiene varios métodos para estimar intervalos de confianza:

- Percentile Bootstrap (método percentil)
- Standard Bootstrap
- Bootstrap-t (“Pivotal Bootstrap”)
- Bias-corrected (BCa) Bootstrap(Sesgo corregido)

Lo importante de esta técnica es que proporciona estimaciones del error estadístico, imponiendo escasas restricciones sobre las variables aleatorias analizadas y estableciéndose como un procedimiento de carácter general, independientemente del estadístico considerado y de las distribuciones de probabilidad de las variables aleatorias consideradas.

Para efectos de nuestro estudio trabajaremos con el método de bootstrap por percentiles y el método estándar de sesgo corregido.

Es necesario recalcar la importancia del muestreo, ya que si la muestra original es una mala representación de la población no importa que técnica de remuestreo utilicen, los resultados llevarán a tomar las decisiones equivocadas.

8.1.2 Fundamentos de la técnica bootstrap

Sea $\theta(F)$ un parámetro dependiente de una función de distribución F desconocida y, por otro lado, las variables aleatorias X_1, X_2, \dots, X_n tales que;

$$X_1, X_2, \dots, X_n \stackrel{iid}{\sim} F$$

Luego, representaremos mediante el conjunto $[X_1, X_2, \dots, X_n]$ una muestra correspondiente a extracciones aleatorias sobre las referidas variables. Note que los valores correspondientes a la muestra nos permiten obtener la distribución empírica \hat{F}_n que constituye la estimación no paramétrica de máxima verosimilitud de la función de distribución F . En apoyo de esta estimación, podemos referirnos al Teorema de Glivenko-Cantelli, que establece una convergencia casi segura, cuando $n \rightarrow \infty$ entre las distribuciones F y \hat{F}_n , aunque debe notarse que es una convergencia asintótica.

$$\sup_x |\hat{F}_n(x) - F(x)| \xrightarrow{c.s} 0$$

Por tanto, es factible establecer $F = \hat{F}_n$ significando que la función de distribución de la variable aleatoria de interés se estima a partir de la distribución empírica, constituyendo el aspecto fundamental del denominado Bootstrap no paramétrico.

En ocasiones, el investigador conoce la función de distribución correspondiente a la variable aleatoria objeto de estudio, aunque se desconozcan los parámetros de la misma, punto que nos conduce al Bootstrap paramétrico.

Supongamos que $\hat{\theta} = \hat{\theta}(x)$ es un estimador de θ , siendo éste un parámetro o vector de parámetros de F , obteniéndose $\hat{\theta}$ a partir de datos muestrales extraídos de la función de distribución F , que depende del parámetro θ , por lo cual puede expresarse $F = F_\theta$. Estimado el parámetro o vector de parámetros, podemos recurrir a la estimación $\hat{F} = F_{\hat{\theta}}$.

El estadístico puede considerar la distribución empírica \hat{F}_n , esta función de distribución empírica es escalonada, o sea, la estimación obtenida de F presenta

discontinuidades, siendo este hecho, precisamente, el punto crítico cuando el estadístico supone la continuidad de F (función suave, derivable en todos los puntos) o posee información adicional que apoya la continuidad de la función de distribución F . En estos casos, es aconsejable suponer que F posee cierto grado de suavidad, o sea, se asemeja parcialmente a una función de distribución derivable en todos sus puntos, que denominaremos \hat{F}^S ; donde s representa las posibles funciones de distribución suaves (por ejemplo: normal, uniforme, etc.). Puede establecerse que la mejor estimación de F proviene de un conjunto de las funciones de distribución F_n y \hat{F}^S , obteniéndose, de esta forma, \hat{F}^C . Este procedimiento de estimación de F se ha denominado bootstrap suavizado.

8.2.1 Estimación del Intervalo de Confianza Bootstrap por Método de los Percentiles.

En este caso, la idea es estimar la función de distribución F del estimador $\hat{\theta}$, usando Bootstrap. Luego, el intervalo de confianza del $100(1-\alpha)$ para θ será simplemente:

$$[\hat{F}^{-1}_{(\frac{\alpha}{2})}, \hat{F}^{-1}_{(1-\frac{\alpha}{2})}]$$

Donde $\hat{F}^{-1}_{(\frac{\alpha}{2})}$ representa el percentil del $\frac{\alpha}{2}$ de la distribución del estimador por bootstrap $\hat{\theta}^*$, es decir, un valor tal que la probabilidad acumulada hasta dicho valor sea $\frac{\alpha}{2}$, y $\hat{F}^{-1}_{(1-\frac{\alpha}{2})}$ representa el percentil del $1 - \frac{\alpha}{2}$.

El procedimiento consiste en generar primero una cantidad B (B aprox 1000) de muestras bootstrap. Luego se calcula en cada una de ellas el estadístico que estima al parámetro y finalmente se calculan los percentiles de la distribución basada en B valores como se ha hecho en los anteriores casos.

8.2.2 Intervalo de Confianza Bootstrap Corregido por Sesgo y Acelerado (BCA).

El intervalo de confianza BCA es una generalización del intervalo de confianza por el método de percentiles, definido para un nivel de confianza del $100(1-\alpha) \%$ por:

$$[\hat{F}^{-1}_{(\frac{\alpha}{2})}, \hat{F}^{-1}_{(1-\frac{\alpha}{2})}] = [\hat{\theta}^{(\frac{\alpha}{2})}, \hat{\theta}^{(1-\frac{\alpha}{2})}]$$

Los extremos del intervalo de confianza BCA también son percentiles de la distribución bootstrap, pero son distintos a los anteriormente mencionados. Los nuevos percentiles dependen de dos números: \hat{z}_0 llamado el factor de corrección por sesgo y \hat{a} llamado el factor de aceleración. Para deducir los valores \hat{z}_0 y \hat{a} hay que considerar el intervalo de confianza del parámetro transformado $f = h(q)$ y donde h es una transformación creciente tal que:

$$\frac{\hat{\phi} - \phi}{se(\phi)} \sim N(-z_0, 1)$$

Donde

$$\hat{\phi} = h(\hat{\theta}) \text{ y } se(\phi) = 1 + a\phi$$

A partir de lo anterior se deduce que el valor de \hat{z}_0 puede ser obtenido directamente de la proporción de valores de las estimaciones de las muestras bootstrap que son menores que el estimado original $\hat{\theta}$. Más específicamente

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#(\hat{\theta} * (b) < \hat{\theta})}{B}\right)$$

Donde Φ^{-1} representa la inversa de la distribución acumulativa de una normal estándar. El valor de \hat{z}_0 puede interpretarse como el sesgo mediana de $\hat{\theta} *$ en unidades normales, es decir, mide la discrepancia entre $\hat{\theta} *$ y $\hat{\theta}$ en unidades normales. Se obtiene un valor $\hat{z}_0 = 0$ si exactamente la mitad de los valores $\hat{\theta} * (b)$ son menores o iguales que $\hat{\theta}$.

El factor de aceleración \hat{a} representa la razón de cambio del error estándar de $\hat{\theta}$ con respecto al verdadero valor del parámetro θ medido en una escala normalizada. La aproximación a la normal $\hat{\theta} \sim N(\theta, se^2)$, asume que el error estándar es el mismo para todo θ , pero esto es algo que raras veces se da y el factor \hat{a} corrige esto. Efron y Tibshirani (1987) proponen usar jackknife para determinar \hat{a} ; Es decir,

$$\hat{a} = \frac{\sum_{i=1}^n (\theta_{jack} - \theta_{(i)})^3}{6(\sum_{i=1}^n (\theta_{jack} - \theta_{(i)})^2)^{3/2}}$$

\hat{a} es 1/6 de la medida de skewness (asimetría) de la función store (derivada parcial de la función de verosimilitud con respecto al parámetro θ).

Finalmente el intervalo de confianza BCA de nivel de confianza $100(1-\alpha)\%$ está dado por

$$\hat{\theta}_{*(\alpha_1)}, \hat{\theta}_{*(\alpha_2)}$$

donde,

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z\left(\frac{\alpha}{2}\right)}{1 - \hat{a}\left(\hat{z}_0 + z\left(\frac{\alpha}{2}\right)\right)}\right)$$

y

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z\left(1 - \frac{\alpha}{2}\right)}{1 - \hat{a}\left(\hat{z}_0 + z\left(1 - \frac{\alpha}{2}\right)\right)}\right)$$

donde Φ representa la función de distribución acumulativa de la normal estándar y $z\left(\frac{\alpha}{2}\right)$ representa el percentil del $100(\alpha/2)\%$ de dicha distribución. Notar que si $\hat{z}_0 = 0$ (estimador insesgado) y $\hat{a} = 0$ (varianza de estimador constante) entonces

$$\alpha_1 = \Phi\left(z\left(\frac{\alpha}{2}\right)\right) = \frac{\alpha}{2} \quad y \quad \alpha_2 = \Phi\left(z\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \frac{\alpha}{2}$$

El intervalo de confianza BCA tiene dos observaciones bastantes importantes:

- 1 Respetar transformaciones del parámetro θ . Esto es si $(\hat{\theta}_{lo}, \hat{\theta}_{up})$ es el intervalo de confianza para θ , entonces $(h(\hat{\theta}_{lo}), h(\hat{\theta}_{up}))$ es el intervalo de confianza para $h(\theta)$.
- 2 Buena precisión. Un intervalo de confianza $(\hat{\theta}_{lo}, \hat{\theta}_{up})$ del $100(1-\alpha)\%$ debe satisfacer que $P(\theta < \hat{\theta}_{lo}) = \frac{\alpha}{2}$ y $P(\theta > \hat{\theta}_{up}) = \frac{\alpha}{2}$. Puede mostrarse que los extremos de un intervalo de confianza BCA tienen precisión de segundo orden; esto es,

$$P(\theta < \hat{\theta}_{lo}) \approx \frac{\alpha}{2} + \frac{c_{lo}}{n} \quad \text{y} \quad P(\theta > \hat{\theta}_{up}) \approx \frac{\alpha}{2} + \frac{c_{up}}{n}$$

- 3 Una debilidad del intervalo de confianza BCA es que puede dar resultados poco precisos para muestras o número de especies pequeñas.

Para solucionar el problema de resultados poco precisos en número de especies pequeñas, estudiaremos el método del percentil, que aproxima mejor los intervalos de confianza.

9 Resultados

Se muestran a continuación, los resultados de la estimación de los intervalos de confianza para los índices de diversidad de Shannon y de Simpson obtenidos mediante remuestreo. Las figuras que siguen a continuación muestran la gráfica de los dos índices considerados y los límites superior e inferior del intervalo de confianza Bootstrap, percentil y sesgo corregido, correspondientes a dichos índices. Recordemos que para la simulación de las distintas poblaciones se utilizaron la distribución normal, la distribución uniforme y la distribución exponencial.

Intervalos de Confianza Bootstrap por el Método del Percentil y el Método Sesgo Corregido BCA para el índice de Simpson y Shannon.

9.1 Índice de Simpson

9.1.1 Población Homogénea

Como ya mencionamos anteriormente, la simulación de los datos para este tipo de población la haremos con la distribución uniforme con parámetros 0 y 1, simulamos con muestras de números de especies entre 10 y 100 cada uno de los estimadores bootstrap del índice de diversidad y su respectivo intervalo de confianza por los métodos de los percentiles y sesgo corregido, en las Figura N°1 y Figura N°2, se muestran los estimadores de los índices y los límites de los intervalos de confianza obtenidos para el índice de Simpson considerando poblaciones homogéneas (distribución uniforme). En las figuras se puede observar un buen comportamiento dentro de las características de los intervalos de confianza bootstrap.

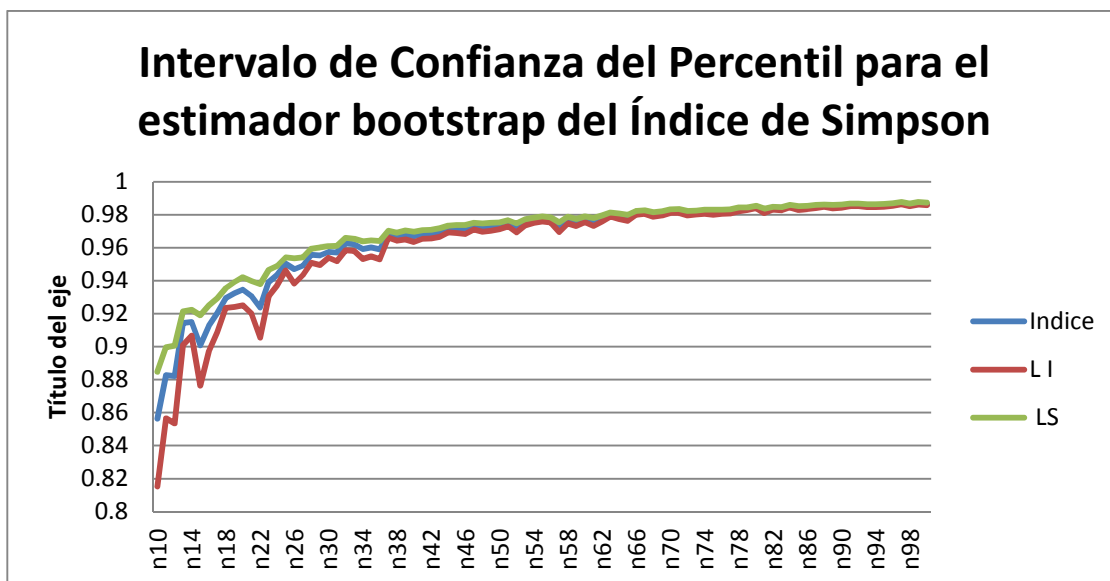


Figura N°1: Intervalos de confianza del percentil para el estimador Bootstrap del Índice de Simpson con distribución Uniforme

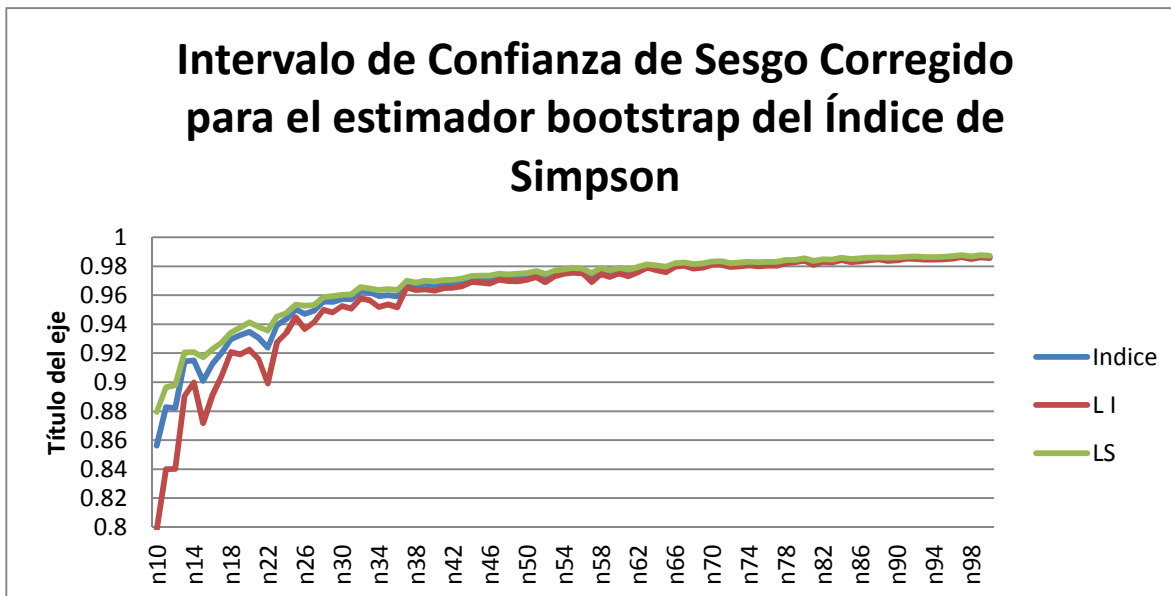


Figura N° 2: Intervalos de confianza de sesgo corregido para el estimador Bootstrap del Índice de Simpson con distribución Uniforme

La figura N°3, presenta una grafica del rango de ambos intervalos de confianza, en ella podemos observar que el intervalo por el método del percentil para valores pequeños de n, número de especies, presenta rangos más pequeños, respecto del intervalo de sesgo corregido, lo cual va cambiando en la medida que n crece y para n suficientemente grande el intervalo de sesgo corregido se hace de rango menor, no obstante en ambos intervalos el rango tiende a cero cuando n es suficientemente grande.

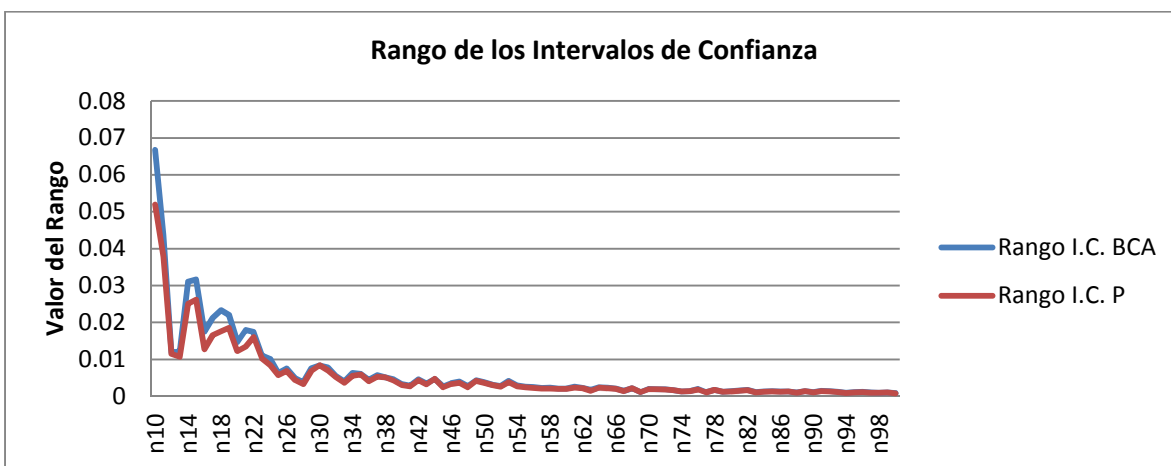


Figura N° 3 Rango de los: Intervalos de confianza para el estimador Bootstrap del Índice de Simpson con Distribución Uniforme

9.1.2 Población con Dominancia

La simulación de los datos para este tipo de población la haremos con la distribución exponencial de parámetro $\lambda = \frac{1}{50}$, simulamos con muestras de números de especies entre 10 y 100 cada uno de los estimadores bootstrap del índice de diversidad y su respectivo intervalo de confianza por los métodos del percentil y sesgo corregido, se muestran los resultados de los estimadores para el índice de Simpson considerando poblaciones con dominancia en los gráficos siguientes.

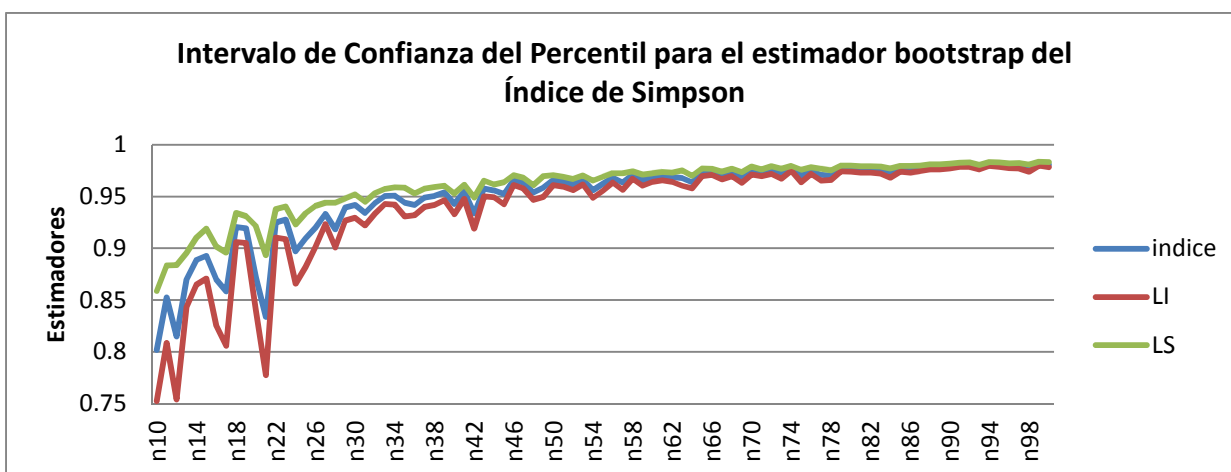


Figura N°4: Intervalos de confianza del percentil para el estimador Bootstrap del Índice de Simpson con distribución exponencial

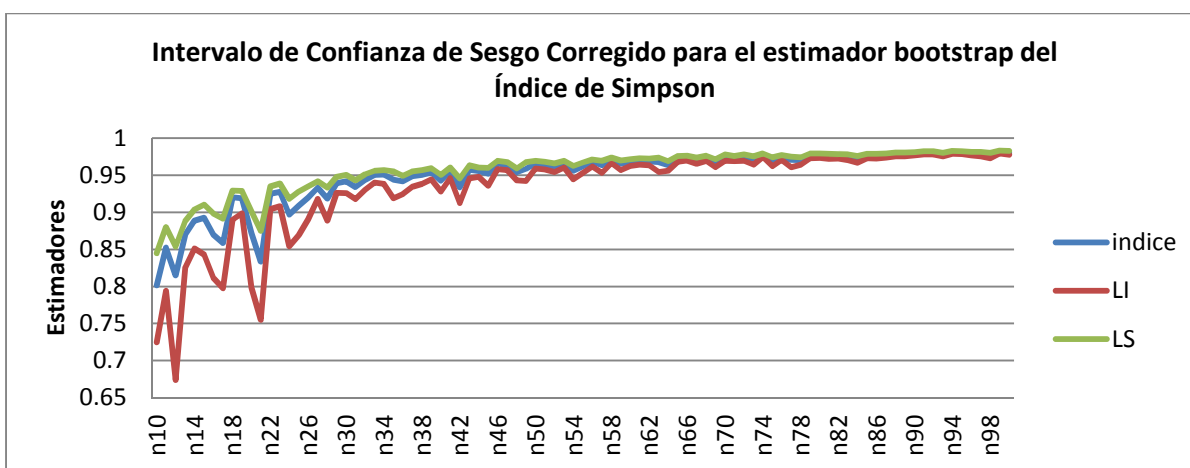


Figura N° 5: Intervalos de confianza de sesgo corregido para el estimador Bootstrap del Índice de Simpson con distribución exponencial.

La figura N°4 y la figura N°5, se presentan las graficas de los estimadores del índice de Simpson y sus intervalos de confianza por el método del percentil (figura

N°4) y por el método de sesgo corregido (figura N°5) en ellos se puede observar mucha similitud entre ambos métodos, aunque para los valores de números de especies pequeños toman valores diferentes el índice y los intervalos pero cuando n crece son muy parecidos los valores del índice y los intervalos para ambos métodos.

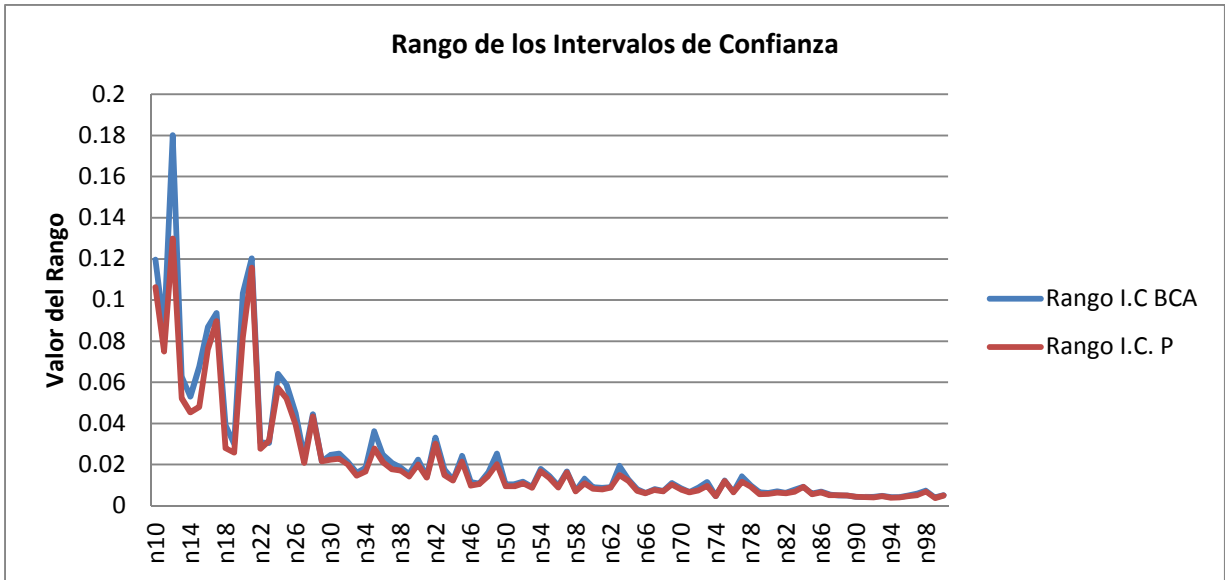


Figura N° 6 Rango de los Intervalos de confianza para el estimador Bootstrap del Índice de Simpson con Distribución Exponencial

En la Figura N°6 podemos observar el comportamiento del rango en los intervalos de percentiles y sesgo corregido para el índice de Simpson en una población con dominancia. Para valores pequeños de n el valor del rango del intervalo por el método del percentil es menor que el rango de sesgo corregido. Esto cambia cuando n tiende al infinito o para valores suficientemente grandes, los rangos se hacen idénticos y su valor cero.

9.1.3 Población Heterogénea

La población heterogénea fue simulada con una distribución normal de parámetros $\mu = 50$ y $\sigma = 25$, simulamos con muestras de números de especies entre 10 y 100 cada uno de los estimadores bootstrap del índice de diversidad y su respectivo intervalo de confianza por los métodos del percentil y sesgo corregido, estos se muestran en los siguientes gráficos.

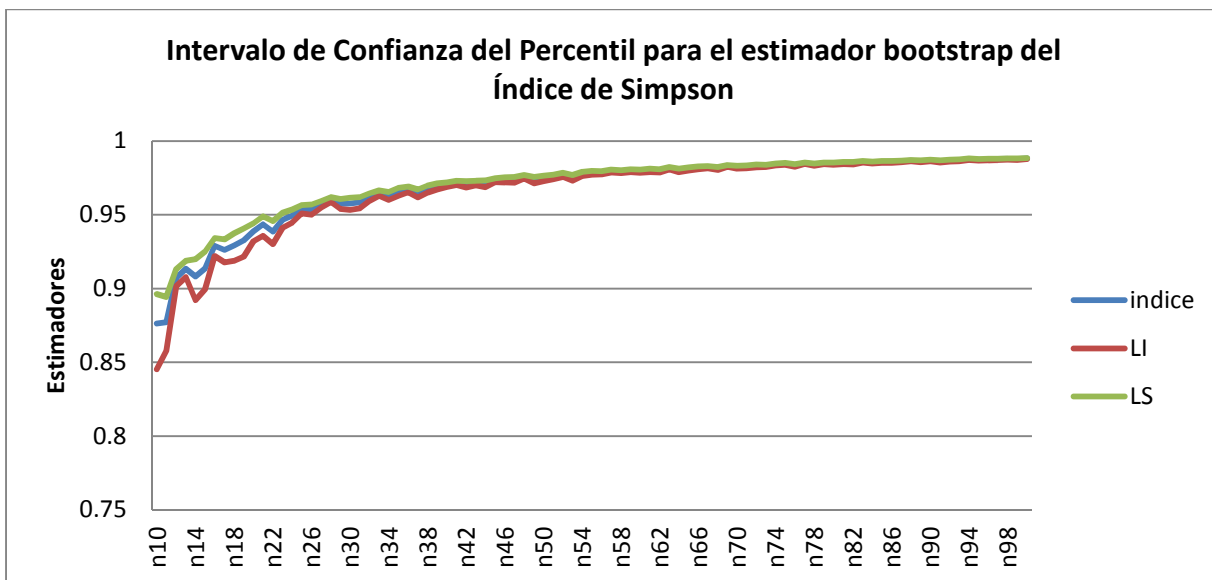


Figura N°7: Intervalos de confianza del percentil para el estimador Bootstrap del Índice de Simpson con distribución normal

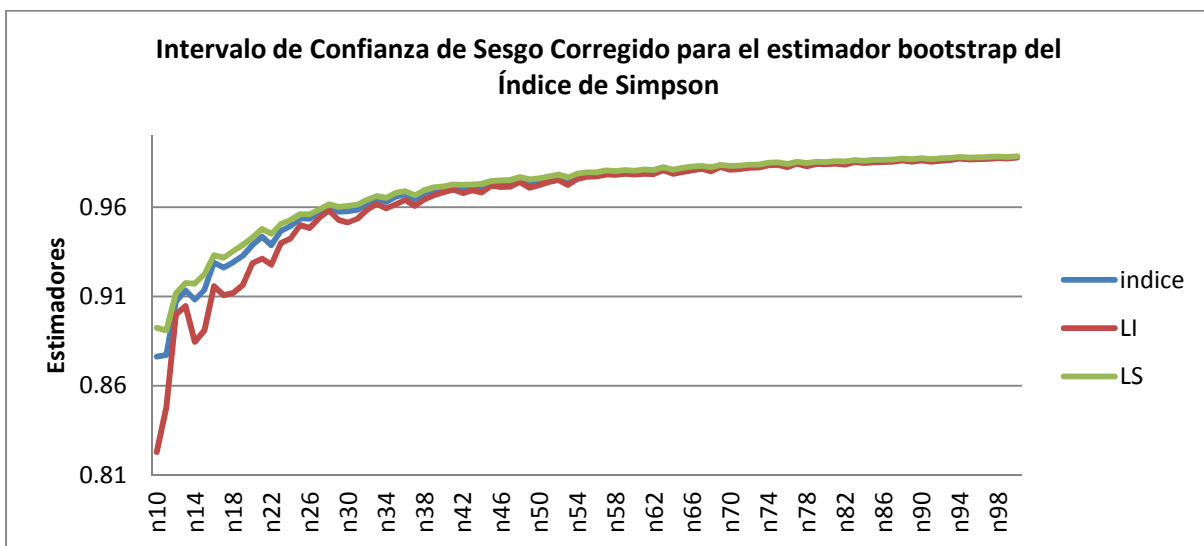


Figura N° 8: Intervalos de confianza de sesgo corregido para el estimador Bootstrap del Índice de Simpson con distribución normal.

En los gráficos anteriores, Figura N°7 y Figura N°8, presentamos el estimador bootstrap para Índice de Simpson y su intervalo de confianza por los métodos del percentil y de sesgo corregido, en ellos podemos observar que el comportamiento presenta una mayor irregularidad para valores pequeños de n principalmente en el intervalo de sesgo corregido, luego si n crece los intervalos son muy similares y no muestran mayor diferencia.

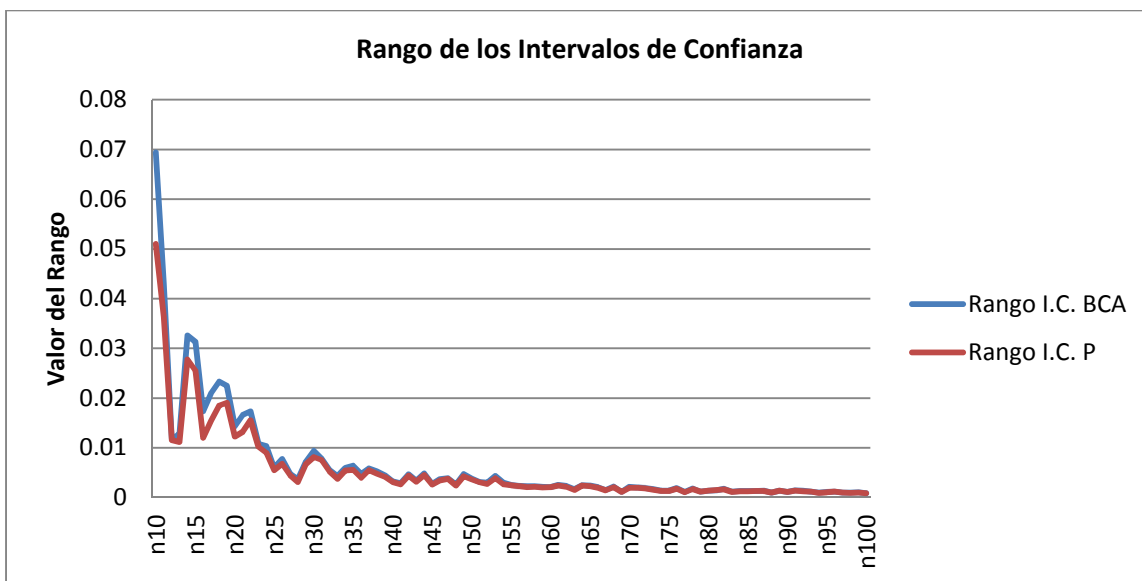


Figura N° 9 Rango de los Intervalos de confianza para el estimador Bootstrap del Índice de Simpson con Distribución Normal

En la Figura N°9, podemos observar la gráfica de los valores del rango de los intervalos por los métodos del percentil y sesgo corregido para el estimador del Índice de Simpson. Como se, el intervalo de sesgo corregido presenta mayor rango en los valores pequeños de n con respecto al rango del intervalo del percentil. Los rangos de ambos intervalos tienden a cero cuando n , número de especies en la muestra, se hace más grande

9.2 Índice de Shannon

9.2.1 Población Homogénea

Como ya mencionamos anteriormente, la simulación de los datos para este tipo de población la haremos con la distribución uniforme con parámetros 0 y 1, simulamos con muestras de números de especies entre 10 y 100 cada uno de los estimadores bootstrap del índice de diversidad y su respectivo intervalo de confianza por los métodos del percentil y sesgo corregido, en las Figura N°10 y Figura N°11, se muestran los estimadores bootstrap para los índices de Shannon y los límites de sus intervalos de confianza obtenidos en poblaciones homogéneas (distribución uniforme). En las figuras se puede observar un buen comportamiento dentro de las características de los intervalos de confianza construidos por los métodos bootstrap.

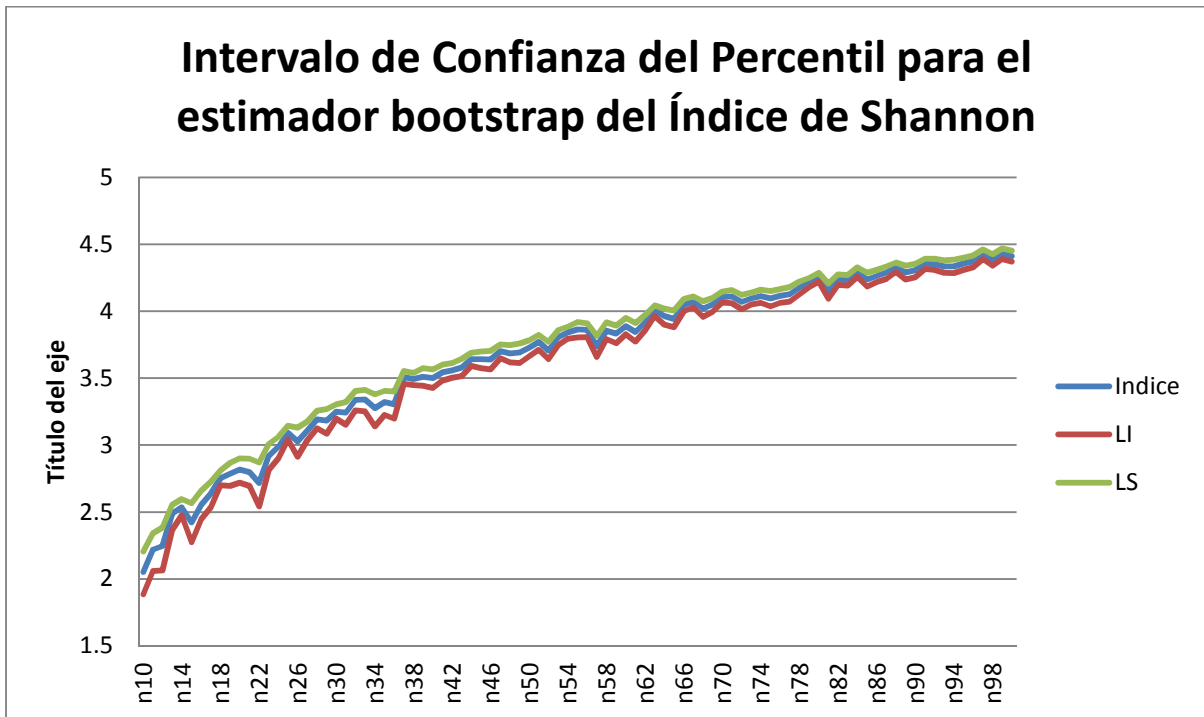


Figura N°10: Intervalos de confianza del percentil para el estimador Bootstrap del Índice de Shannon con distribución Uniforme

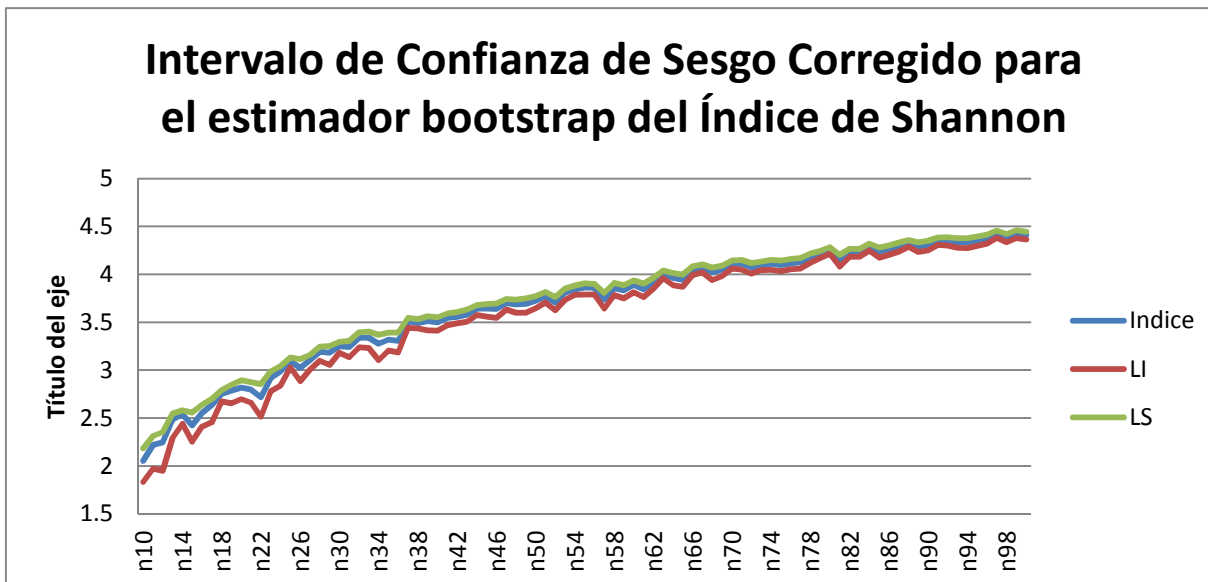


Figura N° 11: Intervalos de confianza de sesgo corregido para el estimador Bootstrap del Índice de Shannon con distribución Uniforme

La figura N°12, presenta una grafica del rango de los intervalos de confianza mostrados en las gráficas anteriores, en las Figura N°10 y N°11, acá se observa que el intervalo del percentil para valores pequeños de n, número de especies, presenta rangos más pequeños y como se ha notado en todos los gráficos anteriores para valores de n pequeños el intervalo del percentil entrega un rango menor lo cual cambia en la medida que n se hace más grande, los rangos toman el mismo valor y además tiende a cero.

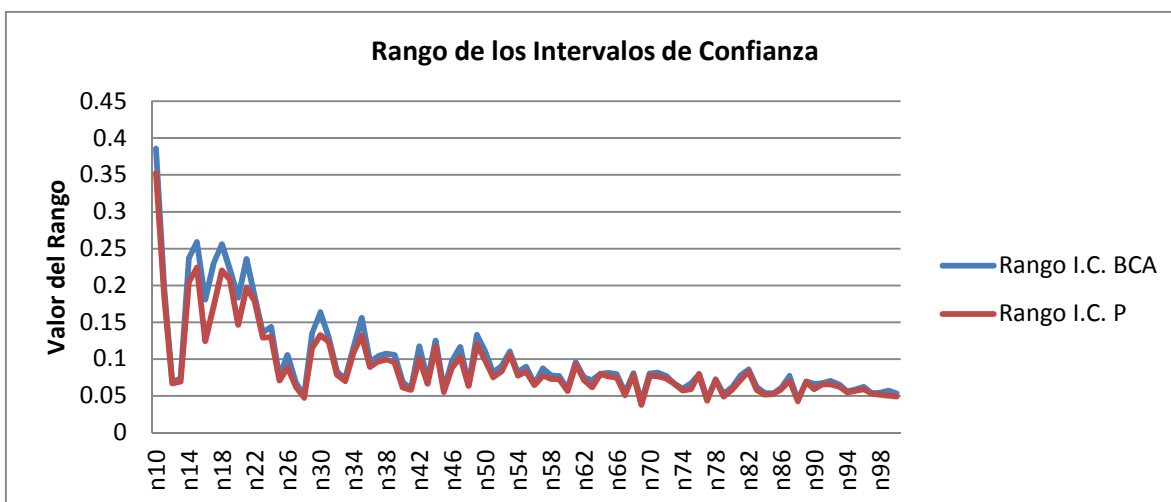


Figura N° 12 Rango de los: Intervalos de confianza para el estimador Bootstrap del Índice de Shannon con Distribución Uniforme

9.2.2 Población con Dominancia

La simulación de los datos para este tipo de población la haremos con la Distribución Exponencial con parámetro $\lambda = \frac{1}{50}$, simulamos con muestras de números de especies entre 10 y 100 cada uno de los estimadores bootstrap del índice de diversidad y su respectivo intervalo de confianza por los métodos del percentil y sesgo corregido, se muestran los estimadores de los índices y los límites de los intervalos de confianza obtenidos para el índice de Shannon considerando poblaciones con dominancia, en los siguientes gráficos.

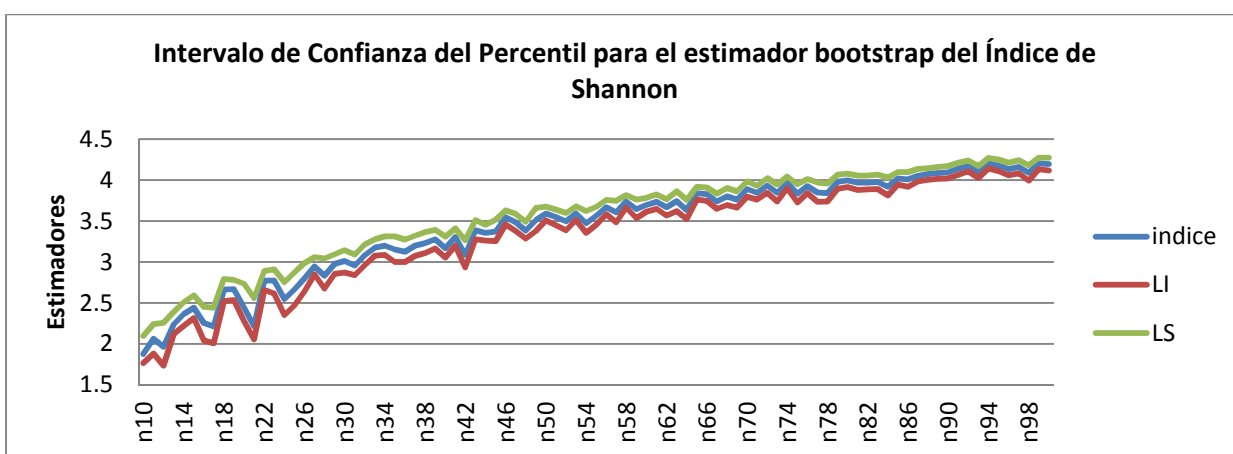


Figura N°13: Intervalos de confianza del percentil para el estimador Bootstrap del Índice de Simpson con distribución exponencial

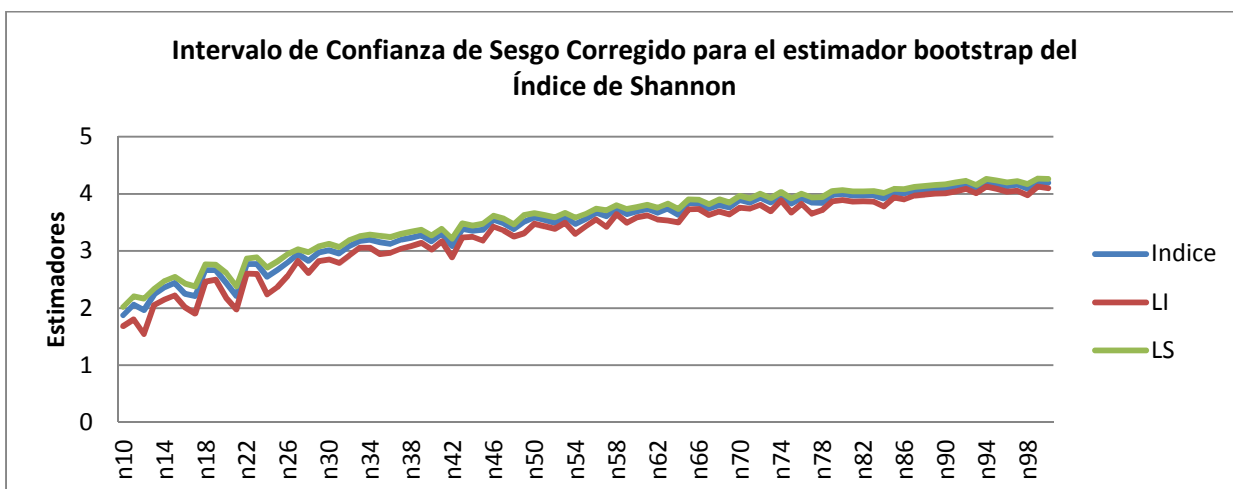


Figura N° 14: Intervalos de confianza de sesgo corregido para el estimador Bootstrap del Índice de Simpson con distribución exponencial.

En las figuras N°14 y N°15 se presentan los intervalos de confianza por métodos bootstrap para el Índice de Shannon, en ellas podemos ver el estimador y los intervalos de confianza contruidos por dos métodos, percentil y sesgo corregido para poblaciones con número de especies normales, es decir heterogéneas, observamos en ellas mucha irregularidad en los valores de n pequeños lo cual se regulariza cuando n crece, además tienen un comportamiento muy similar.

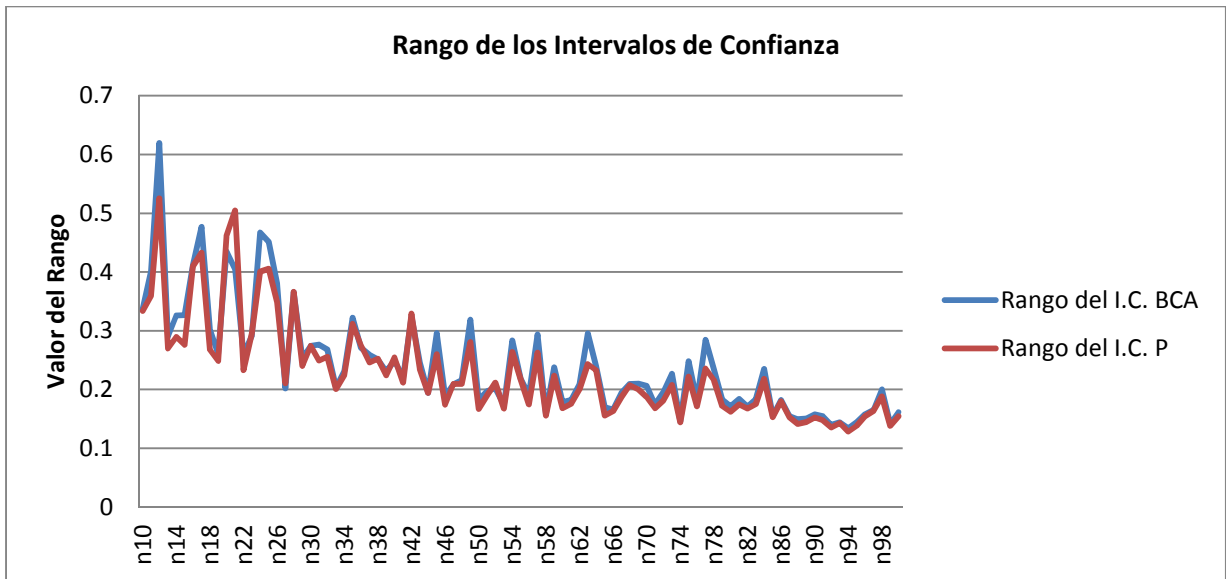


Figura N° 15 Rango de los Intervalos de confianza para el estimador Bootstrap del Índice de Shannon con Distribución Exponencial

La Figura N°15 muestra el valor del rango de los intervalos de confianza de sesgo corregido y del percentil para una población con dominancia. Se observa que el intervalo de sesgo corregido presenta mayores valores que el de percentiles, el valor del rango en ambos intervalos disminuye en la medida que n crece, pero pareciera que más lento que los intervalos anteriores, pero igual el rango se hace cero con n suficientemente grande.

9.2.3 Población Heterogénea

Los datos para una población heterogénea se simularon con la distribución normal para $\mu = 50$ y $\sigma = 25$, simulamos con muestras de números de especies entre 10 y 100 cada uno de los estimadores bootstrap del índice de diversidad y su respectivo intervalo de confianza por los métodos del percentil y sesgo corregido, se muestran los estimadores de los índices y los límites de los intervalos de confianza obtenidos para el índice de Shannon en los siguientes gráficos.

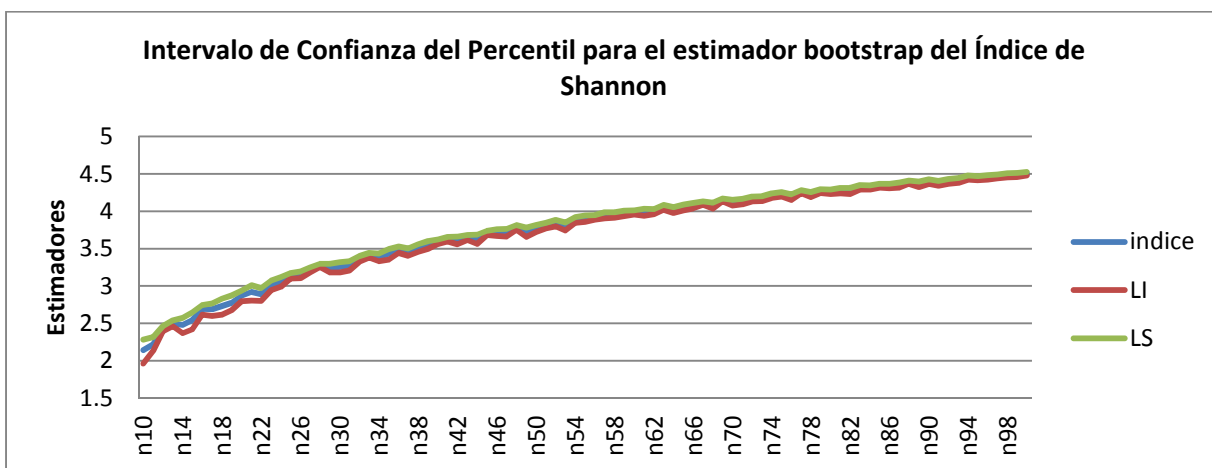


Figura N°16: Intervalos de confianza del percentil para el estimador Bootstrap del Índice de Shannon con distribución normal

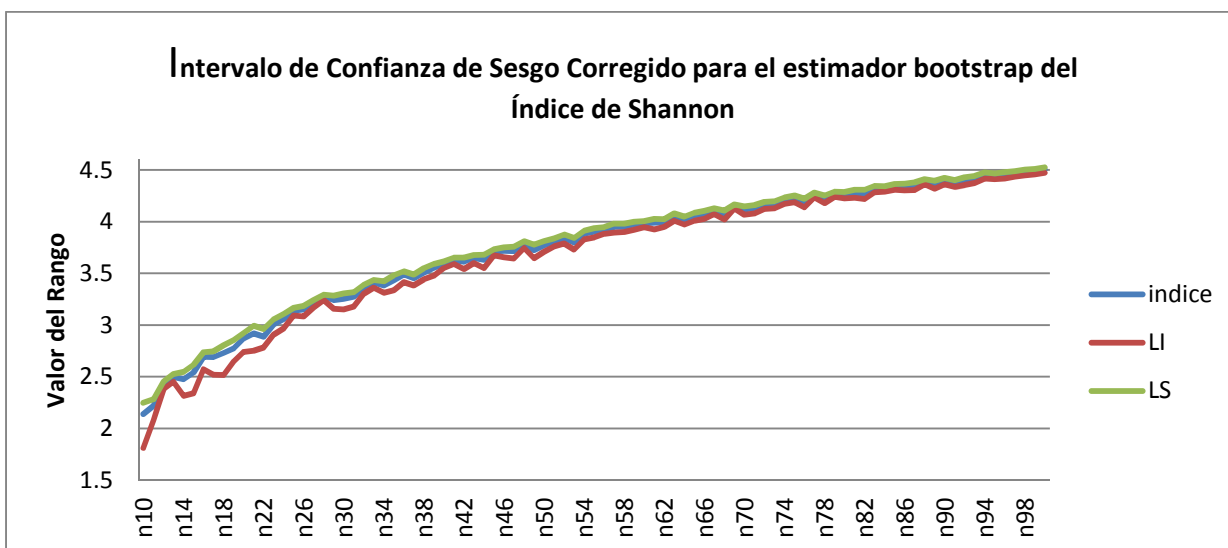


Figura N° 17: Intervalos de confianza de sesgo corregido para el estimador Bootstrap del Índice de Shannon con distribución normal.

Las figuras N°16 y N°17, presentan las gráficas del estimador del índice y el intervalo de confianza para el índice de Shannon de los métodos de sesgo corregido y del percentil, notemos en ellos un comportamiento muy regular de acuerdo a las características de los intervalos bootstrap. Cuando n es pequeño se presenta un poco de irregularidad del intervalo y esto no se nota cuando n es suficientemente grande.

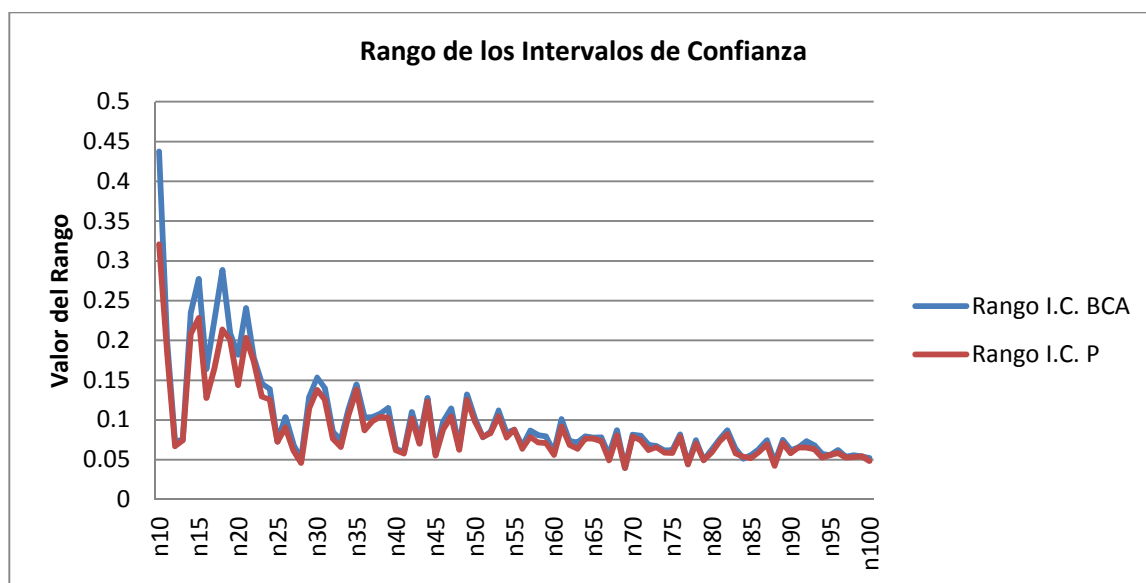


Figura N° 18 Rango de los Intervalos de confianza para el estimador Bootstrap del Índice de Shannon con Distribución Normal

El gráfico, Figura N°18, nos muestra la gráfica del rango de los intervalos de sesgo corregido y del percentil para una población heterogénea del Índice de Shannon. Podemos observar que el rango del intervalo BCA es mayor que el del percentil. Cuando n crece estos valores son similares y además tienden a cero.

9.3 Índices para cada población

Los gráficos de la Figura N° 19 y Figura N° 20, comparan los tres tipos de poblaciones para cada índice.

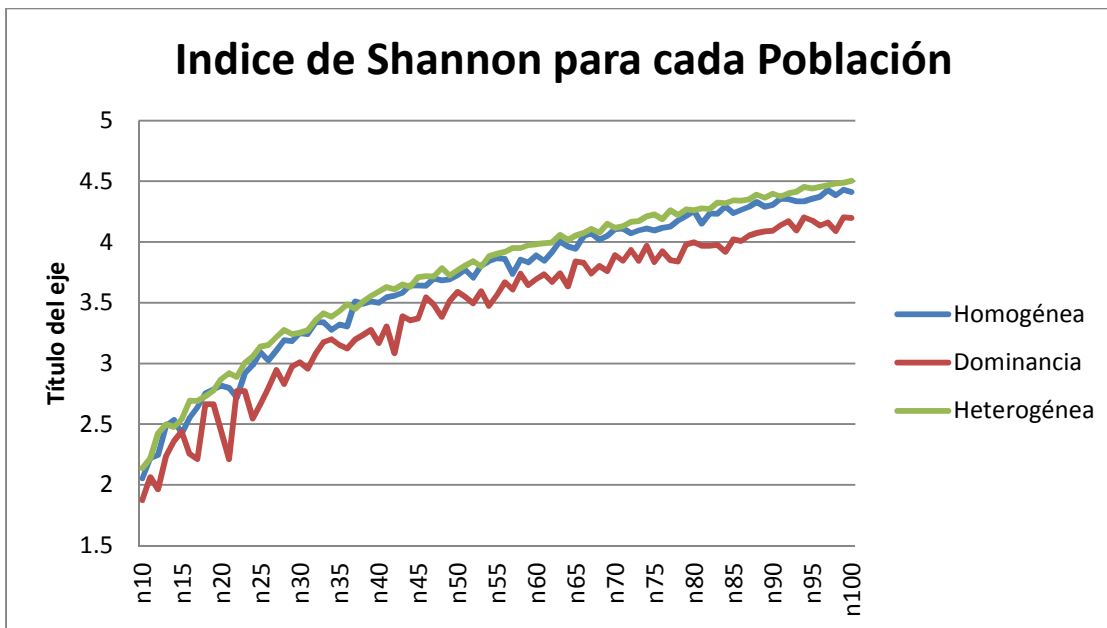


Figura N° 19 Índice de Shannon para cada tipo de Distribución

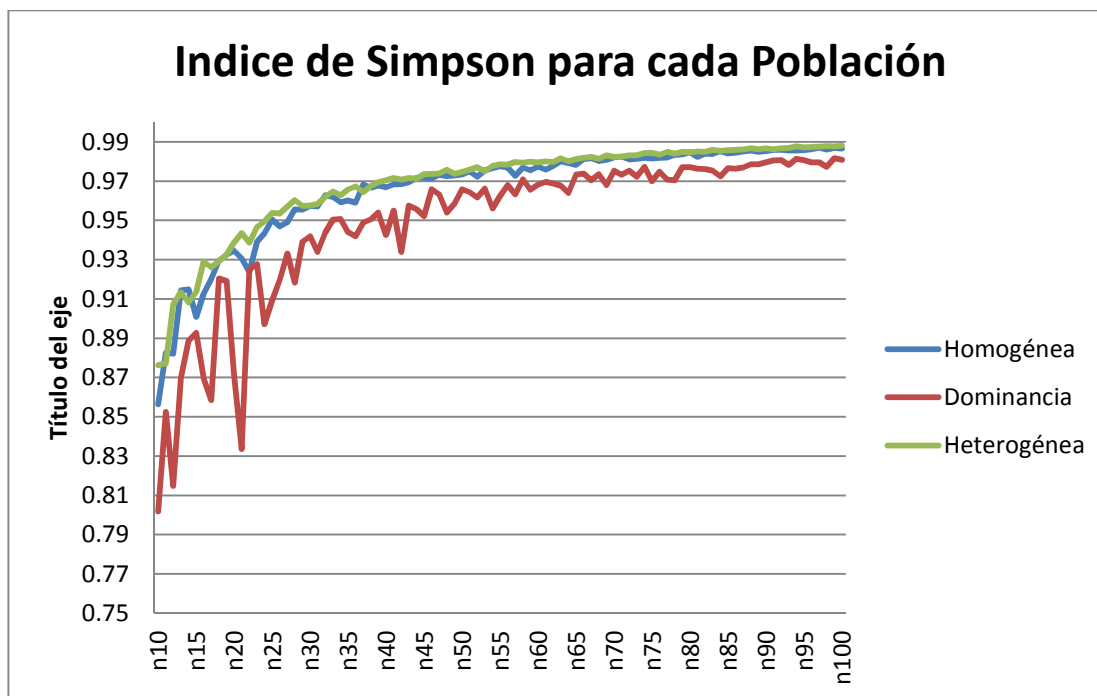


Figura N° 20 Índice de Simpson para cada tipo de Distribución

10 Conclusiones

Como conclusión general respecto de la estimación de estos índices y de la estimación de los intervalos de confianza respectivos, observadas en las Figuras de la 1 a la 18, podemos decir.

En los tres distintos tipos de poblaciones estudiadas, homogéneas, heterogéneas o con dominancia, los intervalos de confianza bootstrap de sesgo corregido y de percentil encontrados muestran comportamientos muy similares, no se percibe una diferencia significativa respecto de sus medidas, en los tres casos el valor de las amplitudes de sus intervalos y de los índices es muy parecido, y no se podría concluir qué método es más eficiente. Está similitud de comportamiento incluye a ambos índices. En efecto, podemos concluir:

10.1 Índice de Simpson.

Los intervalos de confianza para el Índice de Simpson, los cuales se muestran en las Figura N°1 a la Figura N°9. Los métodos del percentil y sesgo corregido, no muestran diferencias en las tres poblaciones (distribuciones) estudiadas. Para los dos métodos de estimación, se tiene que para valores pequeños de n , el número de especies, los intervalos calculados para las tres distribuciones del número de especies (uniforme – asociada a una distribución homogénea de los individuos entre las especies; Normal - asociada a una distribución heterogénea de los individuos entre las especies; Exponencial - asociada a una distribución con dominancia de los individuos entre las especies), presentan un importante nivel de irregularidad. Sin embargo, ésta disminuye en la medida que el número de especies aumenta. De igual manera, la amplitud de los intervalos es mayor cuando n es pequeño, pero, en la medida que aumenta el valor de n , los extremos de los intervalos tienden a valores similares y la amplitud tiende a cero. En poblaciones homogéneas y heterogéneas cuando el número de especies es pequeño, los extremos de los intervalos y de sus índices respectivos, son muy inestables (Fig. 3 y 5), inestabilidad que disminuye cuando el número de especies aumenta.

La amplitud del intervalo presenta el mismo comportamiento. En poblaciones homogéneas y heterogéneas, se puede observar que cuando n es aproximadamente mayor que 50, ésta se estabiliza, tendiendo a cero. En cambio, en poblaciones con dominancia la inestabilidad es mayor y la amplitud se estabiliza aproximadamente cuando el número de especies es mayor a 80; es decir, para el Índice de Simpson, los intervalos de confianza por el método del percentil son más inestables para poblaciones con dominancia y para un menor número de especies.

10.2 Índice de Shannon

Los intervalos de confianza Bootstrap por el método del percentil y sesgo corregido para el Índice de Shannon, representados gráficamente en las Figuras 10 a la Figura 18, no muestran mayores diferencias entre las distribuciones consideradas cuando n crece, respecto del valor de las medidas de los índices y de los intervalos de confianza respectivos. Los intervalos de confianza para poblaciones homogéneas y heterogéneas son muy parecidos a diferencia del intervalo de confianza para población con dominancia que es más irregular.

Este índice es muy inestable para los tres tipos de poblaciones consideradas, para valores de n , número de especies por población, pequeños o grandes. No obstante, se aprecia en las figuras que en la población con dominancia, la inestabilidad es mayor que las poblaciones homogéneas o heterogéneas. La amplitud de los intervalos, en cambio, es mayor para la población heterogénea que para las otras dos, siendo igualmente inestable en todas.

10.3 Comparación de las distribuciones para cada índice.

Las Figuras 19 y 20, muestran los resultados de la estimación de los índices para cada una de las distribuciones de dominancia. La Figura 19, muestra resultados similares del índice de Shannon para las distribuciones uniforme y normal, con un índice de diversidad mayor que para la distribución exponencial, evidenciando que en las distribuciones con mayor dominancia, los índices de diversidad tienden a ser obviamente menores, independientemente del número de especies de la comunidad, pues la diferencia entre éstos estimadores se mantiene uniforme. Del

mismo modo muestra la mayor inestabilidad del índice para la distribución exponencial y un menor número de especies. La Figura 20, muestra, para el índice de Simpson, resultados similares al de Shannon respecto de los valores del índice; sin embargo, muestra también que la inestabilidad de la estimación es mayor que índice de Shannon, para valores menores del número de especies.

Los resultados anteriores, nos permitirían deducir algunas conclusiones claras, respecto que el índice de Shannon es un mejor estimador en el caso de comunidades con menor número de especies.

11 Anexos.

Se detallan a continuación los programas con su comando de salida respectivo para cada índice e Intervalo de Confianza:

11.1 Índice de Simpson.

11.1.2 Intervalo de Confianza del Percentil para poblaciones homogéneas.

```

Bootrat.ic.perc=function(data,B,conf)
{
# Esta funcion calcula el intervalo de confianza
# para un estimado de razon usando Bootstrapping con
# el método de los percentiles
nobs=length(data)
alpha=1-.01*conf
ratio=diversity(data, "simpson")
bootrat=rep(0,B)
for(i in 1:B)
{bootsample=sample(1:nobs,nobs,B)
bootrat[i]=diversity(data[bootsample],"simpson")}
bootrat=sort(bootrat)
kperc=floor((B+1)*alpha*.5)
l.ic=bootrat[kperc]
u.ic=bootrat[B+1-kperc]
cat("El intervalo de confianza por percentiles del",conf,"% para el índice ",ratio,"
es:(",l.ic,",",u.ic,")\n")
}

```

El comando de salida es:

```
Bootrat.ic.perc(Datos,2000,90)
```

Donde

Datos: es la base en donde se almacenan los datos de los números de especies de 10 a 100.

2000: es la cantidad B de remuestreo bootstrap.

90: Nivel de confianza para el intervalo.

11.1.2 Intervalo de Confianza de Sesgo Corregido para poblaciones homogéneas.

```
Bootrat.ic.bca=function(data,conf,B)
```

```
{nobs=length(data)
```

```
alpha=1-.01*conf
```

```
ratio=diversity(data,"simpson")
```

```
ratiob=rep(0,B)
```

```
for(i in 1:B)
```

```
{bootsample=sample(1:nobs,nobs,replace=T)
```

```
ratiob[i]=diversity(data[bootsample],"simpson")
```

```
}
```

```
prop=length(ratiob[ratiob<ratio])/B
```

```
z0=qnorm(prop)
```

```
ratiojackk=rep(0,nobs)
```

```
for(j in 1:nobs)
```

```
ratiojackk[j]=diversity(data[-j],"simpson")
```

```
ratiojkest=mean(ratiojackk)
```



```

desv=ratiojkest-ratiojackk
ahat=sum(desv^3)/(6*(sum(desv^2))^1.5)
z1=qnorm(.5*alpha)
z2=qnorm(1-.5*alpha)
z1c=z0+(z0+z1)/(1-ahat*(z0+z1))
z2c=z0+(z0+z2)/(1-ahat*(z0+z2))
alfa1=pnorm(z1c)
alfa2=pnorm(z2c)
ratiob=sort(ratiob)
kperc1=floor((B+1)*alfa1)
kperc2=floor((B+1)*alfa2)
l.ic=ratiob[kperc1]
u.ic=ratiob[kperc2]

cat("El intervalo de confianza BCA del",conf,"% para el ",ratio," índice
es:(",l.ic,",",u.ic,")\n")
}

```

El comando de salida es:

```
Bootrat.ic.bca(Datos,90,2000)
```

Donde

Datos: es la base en donde se almacenan los datos de los números de especies de 10 a 100.

2000: es la cantidad B de remuestreo bootstrap.

90: Nivel de confianza para el intervalo.

11.1.3 Intervalo de Confianza del Percentil para poblaciones con dominancia.

```

Bootrat.ic.perc=function(data,B,conf)
{
# Esta funcion calcula el intervalo de confianza
# para un estimado de razon usando Bootstrapping con
# el método de los percentiles
nobs=length(data)
alpha=1-.01*conf
ratio=diversity(data, "simpson")
bootrat=rep(0,B)
for(i in 1:B)
{bootsample=sample(1:nobs,nobs,B)
bootrat[i]=diversity(data[bootsample],"simpson")}
bootrat=sort(bootrat)
kperc=floor((B+1)*alpha*.5)
l.ic=bootrat[kperc]
u.ic=bootrat[B+1-kperc]
cat("El intervalo de confianza por percentiles del",conf,"% para el índice ",ratio,"
es:(",l.ic,",",u.ic,")\n")
}

```

El comando de salida es:

```
Bootrat.ic.perc(Datos,2000,90)
```

Donde

Datos: es la base en donde se almacenan los datos de los números de especies de 10 a 100.

2000: es la cantidad B de remuestreo bootstrap.

90: Nivel de confianza para el intervalo.

11.1.4 Intervalo de Confianza de Sesgo Corregido para poblaciones con dominancia.

```

Bootrat.ic.bca=function(data,conf,B)
{nobs=length(data)
alpha=1-.01*conf
ratio=diversity(data,"simpson")
ratiob=rep(0,B)
for(i in 1:B)
{bootsample=sample(1:nobs,nobs,replace=T)
ratiob[i]=diversity(data[bootsample],"simpson")
}
prop=length(ratiob[ratiob<ratio])/B
z0=qnorm(prop)
ratiojackk=rep(0,nobs)
for(j in 1:nobs)
ratiojackk[j]=diversity(data[-j],"simpson")
ratiojkest=mean(ratiojackk)
desv=ratiojkest-ratiojackk
ahat=sum(desv^3)/(6*(sum(desv^2))^1.5)
z1=qnorm(.5*alpha)
z2=qnorm(1-.5*alpha)
z1c=z0+(z0+z1)/(1-ahat*(z0+z1))

```

```

z2c=z0+(z0+z2)/(1-ahat*(z0+z2))
alfa1=pnorm(z1c)
alfa2=pnorm(z2c)
ratiob=sort(ratiob)
kperc1=floor((B+1)*alfa1)
kperc2=floor((B+1)*alfa2)
l.ic=ratiob[kperc1]
u.ic=ratiob[kperc2]

cat("El intervalo de confianza BCA del",conf,"% ",ratio,"para el índice
es:(",l.ic,",",u.ic,")\n")
}

```

El comando de salida es:

```
Bootrat.ic.bca(Datos,90,2000)
```

Donde

Datos: es la base en donde se almacenan los datos de los números de especies de 10 a 100.

2000: es la cantidad B de remuestreo bootstrap.

90: Nivel de confianza para el intervalo.

11.1.5 Intervalo de Confianza del Percentil para poblaciones heterogéneas.

```
Bootrat.ic.perc=function(data,B,conf)
```

```
{
```

```
# Esta funcion calcula el intervalo de confianza
```

```
# para un estimado de razon usando Bootstrapping con
```

```
# el método de los percentiles
```

```
nobs=length(data)
```

```

alpha=1-.01*conf
ratio=diversity(data, "simpson")
bootrat=rep(0,B)
for(i in 1:B)
{bootsample=sample(1:nobs,nobs,B)
bootrat[i]=diversity(data[bootsample],"simpson")}
bootrat=sort(bootrat)
kperc=floor((B+1)*alpha*.5)
l.ic=bootrat[kperc]
u.ic=bootrat[B+1-kperc]
cat("El intervalo de confianza por percentiles del",conf,"% para el índice ",ratio,"
es:(",l.ic,",",u.ic,")\n")
}

```

El comando de salida es:

```
Bootrat.ic.perc(Datos,2000,90)
```

Donde

Datos: es la base en donde se almacenan los datos de los números de especies de 10 a 100.

2000: es la cantidad B de remuestreo bootstrap.

90: Nivel de confianza para el intervalo.

11.1.6 Intervalo de Confianza de Sesgo Corregido para poblaciones heterogéneas.

```
Bootrat.ic.bca=function(data,conf,B)
```

```
{nobs=length(data)
```

```
alpha=1-.01*conf
```

```

ratio=diversity(data,"simpson")
ratiob=rep(0,B)
for(i in 1:B)
{bootsample=sample(1:nobs,nobs,replace=T)
ratiob[i]=diversity(data[bootsample],"simpson")
}
prop=length(ratiob[ratiob<ratio])/B
z0=qnorm(prop)
ratiojackk=rep(0,nobs)
for(j in 1:nobs)
ratiojackk[j]=diversity(data[-j],"simpson")
ratiojkest=mean(ratiojackk)
desv=ratiojkest-ratiojackk
ahat=sum(desv^3)/(6*(sum(desv^2))^1.5)
z1=qnorm(.5*alpha)
z2=qnorm(1-.5*alpha)
z1c=z0+(z0+z1)/(1-ahat*(z0+z1))
z2c=z0+(z0+z2)/(1-ahat*(z0+z2))
alfa1=pnorm(z1c)
alfa2=pnorm(z2c)
ratiob=sort(ratiob)
kperc1=floor((B+1)*alfa1)
kperc2=floor((B+1)*alfa2)
l.ic=ratiob[kperc1]
u.ic=ratiob[kperc2]

```

```
cat("El intervalo de confianza BCA del",conf,"% para el ",ratio," índice
es:(",l.ic,",",u.ic,")\n")
}
```

El comando de salida es:

```
Bootrat.ic.bca(Datos,90,2000)
```

Donde

Datos: es la base en donde se almacenan los datos de los números de especies de 10 a 100.

2000: es la cantidad B de remuestreo bootstrap.

90: Nivel de confianza para el intervalo.

11.2 Índice de Shannon-Wiener.

11.2.2 Intervalo de Confianza del Percentil para poblaciones homogéneas.

```
Bootrat.ic.perc=function(data,B,conf)
```

```
{
# Esta funcion calcula el intervalo de confianza
# para un estimado de razon usando Bootstrapping con
# el método de los percentiles
nobs=length(data)
alpha=1-.01*conf
ratio=diversity(data, "shannon")
bootrat=rep(0,B)
for(i in 1:B)
{bootsample=sample(1:nobs,nobs,B)
bootrat[i]=diversity(data[bootsample],"shannon")}
bootrat=sort(bootrat)
```

```

kperc=floor((B+1)*alpha*.5)
l.ic=bootrat[kperc]
u.ic=bootrat[B+1-kperc]
cat("El intervalo de confianza por percentiles del",conf,"% para el índice ",ratio,"
es:(",l.ic,",",u.ic,")\n")
}

```

El comando de salida es:

```
Bootrat.ic.perc(Datos,2000,90)
```

Donde

Datos: es la base en donde se almacenan los datos de los números de especies de 10 a 100.

2000: es la cantidad B de remuestreo bootstrap.

90: Nivel de confianza para el intervalo.

11.2.2 Intervalo de Confianza de Sesgo Corregido para poblaciones homogéneas.

```

Bootrat.ic.bca=function(data,conf,B)
{nobs=length(data)
alpha=1-.01*conf
ratio=diversity(data,"shannon")
ratiob=rep(0,B)
for(i in 1:B)
{bootsample=sample(1:nobs,nobs,replace=T)
ratiob[i]=diversity(data[bootsample],"shannon")
}
prop=length(ratiob[ratiob<ratio])/B

```



```

z0=qnorm(prop)
ratiojackk=rep(0,nobs)
for(j in 1:nobs)
ratiojackk[j]=diversity(data[-j],"shannon")
ratiojkest=mean(ratiojackk)
desv=ratiojkest-ratiojackk
ahat=sum(desv^3)/(6*(sum(desv^2))^1.5)
z1=qnorm(.5*alpha)
z2=qnorm(1-.5*alpha)
z1c=z0+(z0+z1)/(1-ahat*(z0+z1))
z2c=z0+(z0+z2)/(1-ahat*(z0+z2))
alfa1=pnorm(z1c)
alfa2=pnorm(z2c)
ratiob=sort(ratiob)
kperc1=floor((B+1)*alfa1)
kperc2=floor((B+1)*alfa2)
l.ic=ratiob[kperc1]
u.ic=ratiob[kperc2]

cat("El intervalo de confianza BCA del",conf,"% para ",ratio," el índice
es:(",l.ic,",",u.ic,")\n")
}

```

El comando de salida es:

```
Bootrat.ic.bca(Datos,90,2000)
```

Donde

Datos: es la base en donde se almacenan los datos de los números de especies de 10 a 100.

2000: es la cantidad B de remuestreo bootstrap.

90: Nivel de confianza para el intervalo.

11.2.3 Intervalo de Confianza del Percentil para poblaciones con dominancia.

```

Bootrat.ic.perc=function(data,B,conf)
{
# Esta funcion calcula el intervalo de confianza
# para un estimado de razon usando Bootstrapping con
# el método de los percentiles
nobs=length(data)
alpha=1-.01*conf
ratio=diversity(data, "shannon")
bootrat=rep(0,B)
for(i in 1:B)
{bootsample=sample(1:nobs,nobs,B)
bootrat[i]=diversity(data[bootsample],"shannon")}
bootrat=sort(bootrat)
kperc=floor((B+1)*alpha*.5)
l.ic=bootrat[kperc]
u.ic=bootrat[B+1-kperc]
cat("El intervalo de confianza ",ratio," por percentiles del",conf,"% para el índice
",ratio,"
es:(",l.ic,",",u.ic,")\n")
}

```

El comando de salida es:

```
Bootrat.ic.perc(Datos,2000,90)
```

Donde

Datos: es la base en donde se almacenan los datos de los números de especies de 10 a 100.

2000: es la cantidad B de remuestreo bootstrap.

90: Nivel de confianza para el intervalo.

11.2.4 Intervalo de Confianza de Sesgo Corregido para poblaciones con dominancia.

```
Bootrat.ic.bca=function(data,conf,B)
```

```
{nobs=length(data)
```

```
alpha=1-.01*conf
```

```
ratio=diversity(data,"shannon")
```

```
ratiob=rep(0,B)
```

```
for(i in 1:B)
```

```
{bootsample=sample(1:nobs,nobs,replace=T)
```

```
ratiob[i]=diversity(data[bootsample],"shannon")
```

```
}
```

```
prop=length(ratiob[ratiob<ratio])/B
```

```
z0=qnorm(prop)
```

```
ratiojackk=rep(0,nobs)
```

```
for(j in 1:nobs)
```

```
ratiojackk[j]=diversity(data[-j],"shannon")
```

```
ratiojkest=mean(ratiojackk)
```

```
desv=ratiojkest-ratiojackk
```

```

ahat=sum(desv^3)/(6*(sum(desv^2))^1.5)
z1=qnorm(.5*alpha)
z2=qnorm(1-.5*alpha)
z1c=z0+(z0+z1)/(1-ahat*(z0+z1))
z2c=z0+(z0+z2)/(1-ahat*(z0+z2))
alfa1=pnorm(z1c)
alfa2=pnorm(z2c)
ratiob=sort(ratiob)
kperc1=floor((B+1)*alfa1)
kperc2=floor((B+1)*alfa2)
l.ic=ratiob[kperc1]
u.ic=ratiob[kperc2]
cat("El intervalo de confianza BCA del",ratio,"% para el índice es:(",l.ic,",",u.ic,")\n")
}

```

El comando de salida es:

```
Bostrat.ic.bca(Datos,90,2000)
```

Donde

Datos: es la base en donde se almacenan los datos de los números de especies de 10 a 100.

2000: es la cantidad B de remuestreo bootstrap.

90: Nivel de confianza para el intervalo.

11.2.5 Intervalo de Confianza del Percentil para poblaciones heterogéneas.

```

Bootrat.ic.perc=function(data,B,conf)
{
# Esta funcion calcula el intervalo de confianza
# para un estimado de razon usando Bootstrapping con
# el método de los percentiles
nobs=length(data)
alpha=1-.01*conf
ratio=diversity(data, "shannon")
bootrat=rep(0,B)
for(i in 1:B)
{bootsample=sample(1:nobs,nobs,B)
bootrat[i]=diversity(data[bootsample],"shannon")}
bootrat=sort(bootrat)
kperc=floor((B+1)*alpha*.5)
l.ic=bootrat[kperc]
u.ic=bootrat[B+1-kperc]
cat("El intervalo de confianza por percentiles del",conf,"% para el índice ",ratio,"
es:(",l.ic,",",u.ic,")\n")
}

```

El comando de salida es:

```
Bootrat.ic.perc(Datos,1000,90)
```

Donde

Datos: es la base en donde se almacenan los datos de los números de especies de 10 a 100.

2000: es la cantidad B de remuestreo bootstrap.

90: Nivel de confianza para el intervalo.

11.2.6 Intervalo de Confianza de Sesgo Corregido para poblaciones heterogéneas.

```

Bootrat.ic.bca=function(data,conf,B)
{nobs=length(data)
alpha=1-.01*conf
ratio=diversity(data,"shannon")
ratiob=rep(0,B)
for(i in 1:B)
{bootsample=sample(1:nobs,nobs,replace=T)
ratiob[i]=diversity(data[bootsample],"shannon")
}
prop=length(ratiob[ratiob<ratio])/B
z0=qnorm(prop)
ratiojackk=rep(0,nobs)
for(j in 1:nobs)
ratiojackk[j]=diversity(data[-j],"shannon")
ratiojkest=mean(ratiojackk)
desv=ratiojkest-ratiojackk
ahat=sum(desv^3)/(6*(sum(desv^2))^1.5)
z1=qnorm(.5*alpha)
z2=qnorm(1-.5*alpha)
z1c=z0+(z0+z1)/(1-ahat*(z0+z1))

```

```
z2c=z0+(z0+z2)/(1-ahat*(z0+z2))
alfa1=pnorm(z1c)
alfa2=pnorm(z2c)
ratiob=sort(ratiob)
kperc1=floor((B+1)*alfa1)
kperc2=floor((B+1)*alfa2)
l.ic=ratiob[kperc1]
u.ic=ratiob[kperc2]

cat("El intervalo de confianza BCA del",conf,"% para ",ratio,"el índice
es:(",l.ic,",",u.ic,")\n")
}
```

El comando de salida es:

```
Bootrat.ic.bca(Datos,90,2000)
```

Donde

Datos: es la base en donde se almacenan los datos de los números de especies de 10 a 100.

2000: es la cantidad B de remuestreo bootstrap.

90: Nivel de confianza para el intervalo.

12 Referencias

1. CID, L., 2001. Medidas y estimación de la diversidad. Sustentabilidad de la biodiversidad.../K. Alveal & T. Antezana EDS. Universidad de Concepción- Chile, 29-59.
2. EFRON, B., R. J. TIBSHIRANI. 1987. An Introduction to the Bootstrap. Volumen 57 de Chapman & Hall/CRC Monographs an Statistics & Applied Probability.
3. EFRON, B. 1979. Bootstrap Methods: Another look at the Jackknife. The Annal of Statistics, Vol. 7, N°1 (Jan.,1979), 1-26
4. GASTON, K. J., 1996. Species richness: measure and measurement. In: Biodiversity, a biology of numbers and difference. K. J. Gaston (Ed.) Blackwell Science, Cambridge, 77-113
5. HURLBERT, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological monographs, **54**: pp. 187-211.
6. MALLET, J. 1996. The genetics of biological diversity: from varieties to species. In: Biodiversity: a biology of numbers and difference, K. J. Gaston (Ed) Oxford University Press, Oxford, 41-57.
7. MAGURRAN, A. E. 1988. Ecological diversity and its measurement. Princeton University Press, New Jersey, pp. 179.
8. PEET, R. K, 1974. Atlas de Biodiversidad Marina del Mar Balear, <http://www.ba.ieo.es/bioatlasmarino/indices-de-biodiversidad>
9. SHAPOSHNIKOV, M 1966. El Síndrome de Asperger, la Ciencia y Yo <http://aspergeryciencia.com/algo-sobre-la-ciencia/>
10. SMITH, E.P, 2002. Estimación del Índice de Diversidad de Simpson en m Sitios de Muestreo. <http://rev-inv-ope.univ-paris1.fr/files/26205/IO-26205-9.pdf>
11. SOLBRIG, O. T., 1991. From genes to ecosystems: a research agenda for biodiversity. IUBS, París.

12. SPELLERBERG, I. F., 1991. *Monitoring ecological change*. Cambridge University Press, UK, 334.
13. UNEP, 1992. Convention on biological diversity. United Nations Environmental Program, Environmental Law and Institutions Program Activity Centre. Nairobi.
14. WHITTAKER, R. H., 1972. Evolution and measurement of species diversity. *Taxon*, 21, 213-251.
15. WILSON, E.O, 1988. www.uco.es/servicios/diversidad.ppt/index.php(láminas N°11 a N°16)