

# ANÁLISIS ESTADÍSTICO DE SISTEMAS ECOLÓGICOS



---

**LUIS MIGUEL MEJÍA GIRALDO**



Afiliada a la Asociación Colombiana de Universidades "ASCUN"

# ANÁLISIS ESTADÍSTICO DE SISTEMAS ECOLÓGICOS

---

Investigador Principal  
**LUIS MIGUEL MEJÍA GIRALDO**

**UNIVERSIDAD LA GRAN COLOMBIA**  
**Seccional Armenia**

**ANÁLISIS ESTADÍSTICO DE SISTEMAS ECOLÓGICOS**

Autor:  
LUIS MIGUEL MEJÍA GIRALDO

ISBN: 978-958-8510-12-5  
Ejemplares: 100  
© Todos los Derechos Reservados  
Año 2009

Revisión de Estilo: Juan Manuel Acevedo Carvajal  
Diseño Carátula: Optigraf S.A.  
Diagramación: Optigraf S.A.  
Impresión: Optigraf S.A.

## **HONORABLE CONSILIATURA**

**Germán Darío Ledesma López**  
Presidente

**José Galat Noumer**  
Rector General

**Carlos Alberto Pulido Barrantes**  
Secretario

**Raúl Abril Cárdenas**  
Consiliario

**Rafael Diazgranados Peñaranda**  
Consiliario

**Eduardo Carvajalino Contreras**  
Consiliario

**Roberto Herrera Soto**  
Consiliario

**María Consuelo Castaño Triana**  
Consiliaria

**Teodoro Gómez Gómez**  
Representante de los Profesores

**Jeimy Cadena Duque**  
Representante de los Estudiantes

**Myriam Luz Vargas**  
Revisora Fiscal

**UNIVERSIDAD  
LA GRAN COLOMBIA**  
Armenia

## **AUTORIDADES UNIVERSITARIAS**

**José Galat Noumer**  
Rector General

**Jaime Bejarano Alzate**  
Rector delegatario

**Bibiana Vélez Medina**  
Vicerrectora Académica

**Jorge Alberto Quintero Pinilla**  
Vicerrector Administrativo y Financiero

**Ana Milena Londoño Palacio**  
Secretaria General

## ANÁLISIS ESTADÍSTICO DE SISTEMAS ECOLÓGICOS

**LUIS MIGUEL MEJÍA GIRALDO**

Investigador Principal

*Ingeniero Agrónomo de la Universidad de Caldas, Especialista en Gestión para el Desarrollo Empresarial de la Universidad Santo Tomás y Especialista en Pedagogía y Docencia Universitaria de la Universidad La Gran Colombia.*

## TABLA DE CONTENIDO

	Pág.
Introducción .....	9
<b>1. Manejo de datos en comunidades ecológicas .....</b>	<b>11</b>
Aspectos Asociados a Variables e indicadores .....	15
Criterios útiles para la Selección de Variables .....	16
-Método Florístico .....	17
-Método Fisionómico o No Florístico .....	18
Escalas de Registro .....	18
-Densidad .....	18
-Cobertura .....	20
- Area Basal .....	20
-Biomasa .....	21
-Frecuencia .....	21
-VARIABLES Combinadas .....	21
<b>2. Parámetros y Aplicaciones del Muestreo para Sistemas Ecológicos .....</b>	<b>23</b>
Antecedentes .....	23
Muestreo Dirigido, no Aleatorio .....	24
Muestreo Aleatorio simple .....	24
Muestreo Aleatorio Estratificado .....	30
Muestreo Sistemático o Secuencial .....	32
Muestreo por Conglomerados .....	34
Muestreo Anidado .....	35
<b>3. Análisis de patrón espacial .....</b>	<b>41</b>
Antecedentes .....	41
Métodos de distribución .....	42
Métodos de varianza-cuadrante .....	49
Métodos de distancia .....	55

<b>4.Relaciones de abundancia de especies .....</b>	<b>62</b>
Antecedentes .....	62
Modelos de distribución .....	62
Índices de diversidad  .....	72
Índices de Uniformidad .....	76
<b>5. Afinidad de especies .....</b>	<b>81</b>
Antecedentes .....	81
Índices de traslape de nichos .....	81
Asociación interespecífica .....	89
Covariación interespecífica .....	102
<b>6.Clasificación de comunidades.....</b>	<b>108</b>
Antecedentes .....	108
Funciones de semejanza .....	109
Análisis de asociación .....	116
Análisis de clusters .....	120
<b>Bibliografía .....</b>	<b>126</b>

***Dedicatoria,***

*Este libro está dedicado a mis hijos Santiago y Juliana, a mi esposa Liliana María y a mis estudiantes porque he tenido la fortuna de aprender a su lado.*

*Ubi Dubium Ibi Libertas (Donde hay duda, hay libertad).  
Proverbio Latino*



## INTRODUCCIÓN

En la actualidad, la innovación tecnológica y la Investigación y Desarrollo (I+D) están basadas en herramientas estadísticas estructuradas para, no sólo analizar, sino también para apoyar científicamente la elaboración nuevos productos u optimizar procesos, como estrategias fundamentales en el crecimiento y desarrollo de regiones y países, donde se debe entender los recursos renovables que nos circundan con una mirada de conversión de ventaja comparativa a ventaja competitiva y para ello se requieren una serie de herramientas y entre ellas las de carácter estadístico encaminadas hacia la investigación y la optimización de procesos agrícolas y agroindustriales soportados en una adecuada comprensión de nuestros sistemas ecológicos.

Ya que, con una visión de investigador y de equipo interdisciplinario de investigadores es necesario estudiar el comportamiento de sistemas ya sean estos abiertos o cerrados, bajo unas condiciones y pruebas, en algunos casos deductivas, en otros experimentales, resaltándose entonces la necesidad de la aplicación de métodos de evaluación para determinar el comportamiento de comunidades bióticas, apuntando hacia la generación de información pertinente con respecto a aspectos propios de los procesos ecológicos bajo estudio con miras a un adecuado manejo agroindustrial de ser posible.

Además, en la aplicación y difusión de métodos estadísticos en la investigación, a través de un conjunto de técnicas existentes que se obtienen nuevos avances de un proceso o sistema en particular y poder inferir mediante experimentación el comportamiento de un conjunto de variables que se tengan y la posible diferencia estadística entre ellas.

El análisis estadístico de ecosistemas ha llegado para posicionarse como base para adoptar nuevas técnicas que propenden por el mejoramiento de la calidad de vida; recordando que el propósito de la ciencia estadística es suministrar una base objetiva para el análisis de problemas en los que la información con la que se cuente se aparte de la causalidad exacta.

Es por lo anteriormente expuesto que se ha ideado un sistema lógico general de razonamiento inductivo y deductivo aplicable a datos de esta naturaleza. A su vez; la estadística en su contexto es aplicable en diferentes áreas como control de calidad, procesos productivos, bacteriología, análisis de índole agrícola y pecuario, biotecnología, entre otros, y es por lo anteriormente expuesto que se requiere preparar investigadores que apliquen con el criterio adecuado tal herramienta estadística, dado que su certeza radica en la selección del diseño a partir de la naturaleza de lo que se va investigar.

No se pueden dejar pasar los aportes dados por Sir Ronald Fisher (1890-1962), Karl Pearson (1857-1936), Jerzy Neyman (1894-1981), por mencionar solo unos cuantos científicos a quienes les debemos lo que la estadística es hoy y que gracias a ellos la investigación estadística será un mundo de exploración y encantamiento al permitirnos "leer" y empezar a entender nuestro entorno y como en el caso del presente libro, el entorno ecológico, lo cual es un comienzo para el entendimiento de muchos más, siendo así, intelectualmente estimulante.

## 1. MANEJO DE DATOS EN COMUNIDADES ECOLÓGICAS.

Cuando se lleva a cabo una exploración en comunidades bióticas (bosques, lagos, selvas, reservas forestales, entre otros) se presentan diferentes patrones de comportamiento como dispersión de especies, relación, al interior de comunidades poblacionales.

No obstante, se deben diferenciar los datos tomados, ya que estos se clasifican en experimentales y observacionales, los cuales presentan las siguientes definiciones:

- **Experimental:** Aquel donde se puede dividir una comunidad en porciones replicadas en varios tratamientos y controles que pueden ser impuestos por el investigador.
- **Observacional:** Cuando se hacen medidas a las comunidades bióticas en un rango de condiciones impuestas por la naturaleza más que por el investigador.

Es con este tipo de toma de datos que:

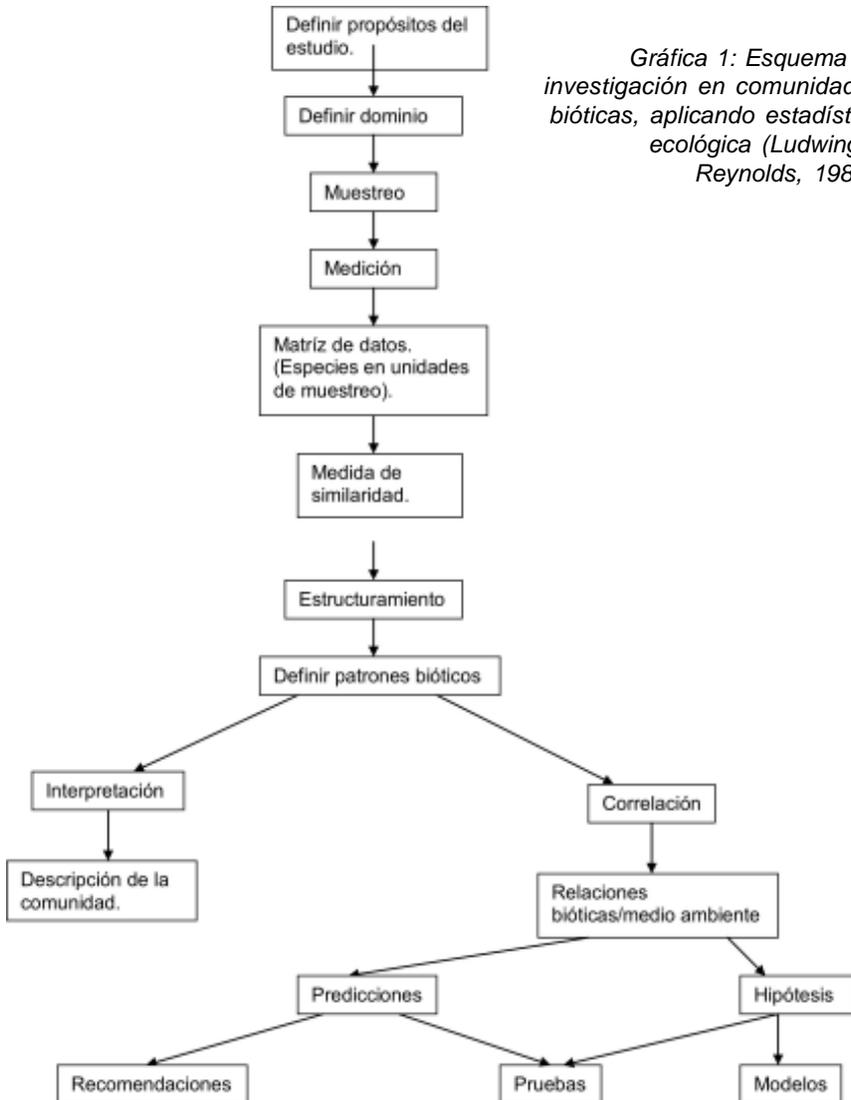
- Se estudian diferentes muestras obtenidas al mismo tiempo en diferentes condiciones (posible variación en el espacio).
- Se estudian diferentes muestras en el mismo lugar pero en diferentes tiempos (posible variación en el tiempo).

Cuando se toman los datos de tipo observacional, la información alcanzada está asociada a la variabilidad dentro de una comunidad sin ser impuesta; además, la aproximación observacional puede ser de índole inductivo, multivariado y no experimental.

Como se mencionó anteriormente, la toma de datos observacionales difieren en el tiempo o espacio, pudiéndose determinar, según las necesidades del investigador la composición de especies en un área, propiedades de correlación de especies en función de los factores ambientales circundantes, detección de patrones de comportamiento específicos, entre otros tipos de análisis.

Sin embargo, es de resaltar la importancia del muestreo ecológico, ya que el valor de los datos obtenidos depende de la exactitud y precisión del mismo procedimiento.

Los estudios e investigaciones de análisis ecológicos que requieren el uso de herramientas estadísticas manejan un esquema básico y flexible (ver gráfica 1).



Gráfica 1: Esquema de investigación en comunidades bióticas, aplicando estadística ecológica (Ludwig y Reynolds, 1988).

Como se observa en la gráfica, cuando se lleva a cabo un estudio ecológico utilizando herramientas estadísticas deben ser claros los propósitos del estudio, que es?, para que sirve?, porqué?, como?, cuando? y demás cuestionamientos que permitan profundizar en dichos propósitos.

Seguido a lo anterior se define el dominio (sitio de colección de datos, la comunidad biótica a evaluar en este caso), se realiza el muestreo y la respectiva medición de los datos, los cuales se consignan en tablas sencillas para el posterior análisis, conocidas también como sábanas, ya sean estas función del tiempo (ver tabla 1) o espacio (ver tabla 2).

Así como se observa en la gráfica 1, se procesa la matriz de datos (de las especies en las unidades de muestreo evaluadas en la colección de datos), con base en medidas de similaridad, que son las diferentes mediciones que se toman, no sólo de las especies bajo investigación, sino también de las condiciones ambientales reinantes como temperatura, brillo solar, humedad relativa, presión atmosférica y altitud. De esta manera se logra el estructuramiento del análisis de estadística ecológica para el estudio de una comunidad biótica dada.

Después de pasar por las fases colección de datos y análisis de los mismos, se definen los patrones bióticos, basados en el cuestionamiento inicial al definir el propósito del estudio, los cuales pueden ser de patrón espacial, relaciones de abundancia de especies, afinidad de especies, clasificación de comunidades, entre otros patrones de índole biótico. También, se pueden establecer sistemas de monitoreo al interior de las comunidades de las especies en cuestión.

Seguidamente, existen dos caminos a tomar, el primero es la descripción de la comunidad biótica evaluada, mientras que el segundo se centra en las relaciones bióticas en interacción con el medio ambiente que circunda a la comunidad bajo evaluación, orientadas hacia la predicción de comportamientos o hipótesis acerca de dicha comunidad, estableciéndose así recomendaciones, pruebas y modelos.

Es de resaltar que, es de suma importancia un esquema adecuado de muestreo, el cual debe estar basado en unidades de muestreo (UM) apropiadas según las condiciones de la investigación planteada, como son hojas de una planta, trampas de luz, sitios de cateo, organismos individuales, trampas de fosa y cinturones transectos (bajo una selección natural o arbitraria); mientras que una muestra es la colección de tales unidades de muestreo.

Además, para llevar a cabo un adecuado procedimiento de muestreo siempre se deben definir, inicialmente, las unidades (UM) y las medidas específicas como presencia/ausencia de organismos, biomasa, densidad, incidencia, intensidad, entre otras, para alcanzar una satisfactoria recolección de los datos.

**Tabla 1: tabla para la recolección de datos de comunidades ecológicas en función del tiempo (t).**

	Unidades de muestreo (UM) en función del tiempo t				
Especie	1	2	3	. . .	t
A					
B					
C					
. . .					
Factores ambientales					
W					
X					
Y					
Z					

**Tabla 2: tabla para la recolección de datos de comunidades ecológicas en función del espacio (n).**

	Unidades de muestreo (UM) en función del espacio n				
Especie	1	2	3	. . .	n
A					
B					
C					
. . .					
Factores ambientales					
W					
X					
Y					
Z					

● **Aspectos asociados a Variables e Indicadores.**

Tales medidas se basan en la toma de variables e indicadores, que son cifras o datos configurados de manera científica a través de la operacionalización metodológica, con el fin de que reflejen el comportamiento de aquello que se piensa investigar, generando en última instancia información útil para la toma de decisiones.

Las variables son características o cualidades asociadas a la realidad que poseen la capacidad de asumir valores y por ende pueden variar. Las variables se clasifican en:

- Cualitativas: no se pueden medir, pero son, como su nombre lo indica, determinantes de cualidades y por ello se clasifican a su vez en nominales (femenino o masculino, presencia o ausencia) y ordinales (Grados: Alto, medio, bajo).
- Cuantitativas: variables que se pueden medir y se clasifican en continuas (permiten decimales como temperatura, peso, altura,

proporciones) y discretas (variables de conteo como número de insectos por trampa, número de plantas por hectárea).

Cuando deseamos analizar variables que determinan fenómenos como rendimiento, calidad, diversidad, debemos recurrir a la descomposición de tales variables en cualidades asociadas al fenómeno de interés, recurriendo así a los indicadores.

Los indicadores son componentes de las variables que poseen relativa autonomía y la asociación de un grupo bien determinado de ellos estructura las variables.

Por lo anteriormente expuesto, el uso de variables e indicadores requieren de la correcta aplicación de metodologías que los validen. No obstante, es de notar que las condiciones de los ecosistemas hacen que estos sean dinámicos y requieren de una elección de técnicas que dependerán de variables fiables y válidas y que su toma posea características de trazabilidad que se basa en la calidad de la información en función de la exactitud y la precisión, dado que es de suma delicadeza la validación y manipulación de datos obtenidos y requieran técnicas apropiadas de recolección, de calibración de equipos si son requeridos y de almacenamiento, como se pudo observar en los ejemplos de sábanas (Tablas 1 y 2). De igual manera, dichas técnicas dependen de los costos económicos, del tiempo disponible para el estudio, pero el recurso más importante es el conocimiento previo que se tenga de la zona a estudiar.

- ***Criterios útiles para la selección de variables.***

Cuando se van a seleccionar variables para estudios de sistemas ecológicos, se deben preferir aquellas que cumplan con las necesidades de la investigación pero que a su vez sean de menor costo económico, tiempo y mayor simplicidad y para determinar esto los objetivos deben estar bien estructurados e identificados.

El hecho es tomar aquellas variables que expliquen el fenómeno o fenómenos de interés para el investigador y no todas las variables.

En los estudios dirigidos hacia comunidades bióticas es necesario saber ¿Qué seres vivos habitan un ecosistema?, ¿Cuántos hay?, ¿Cómo se distribuyen?, ¿Cómo cohabitan?, ¿Cómo cambia su estructura en función del tiempo y espacio?, entre otros cuestionamientos y para resolverlos se debe definir de manera sistemática las variables e indicadores a evaluar, según el propósito del estudio.

Sin embargo, las comunidades pueden variar considerablemente en extensión y la simple inclusión de un listado de especies no determina el comportamiento de ellas y mucho menos su comportamiento en el entorno que las circundan.

En cuanto a la elección del método para describir vegetación, ésta depende de varios factores tales como el propósito del estudio, atributos asociados a las plantas y animales, escala geográfica, tipo de hábitat y la necesidad de identificar o no las especies.

En cuanto a fauna refiere, las técnicas para su estudio contemplan la detección directa de los individuos, ya sea por avistamiento, captura, restos de animales o por estimadores indirectas basadas en indicadores de presencia o actividad como son huellas, fecas, nidos, daños a plantas, entre otros.

Según Graf y Sayagués (2000) la caracterización de la vegetación se realiza en función de una serie de categorías en que se registran las plantas de la comunidad y se basan en 2 métodos: florístico y fisionómico (conocido también como no florístico), respectivamente.

El método florístico describe la flora en su totalidad, mientras que el método fisionómico (no florístico) describe la estructura de la vegetación en términos generales.

### **1. Método Florístico:**

Describe la flora en su totalidad al interior de una caracterización y se lleva a cabo estableciendo la composición botánica de las especies que la integran.

## **2. Método Fisionómico o No Florístico:**

Las plantas son analizadas en función de su morfología, fenología, valor productivo u otros criterios no florísticos.

Sin embargo, este método posee múltiples sistemas de clasificación de plantas con diferentes grados de elaboración (Graf y Sayagués, 2000).

Al hablar de fisionomía (Aspecto), esta se define por las diferentes formas de crecimiento presentes (tales como árboles, arbustos, etc), su estratificación, grado de cobertura y características del follaje.

Otros criterios son ciclo de vida (anual, perenne) y hábito de crecimiento (por estolón, rizomas, cespitoso, maciego).

### **● Escalas de Registro:**

Para el análisis de sistemas ecológicos es de suma importancia elegir las escalas de registro que son de tipo nominal, ordinal y proporción.

Dentro de las variables más popularizadas están:

**1. Densidad:** Esta variable se basa en el número de plantas por unidad de área. La densidad puede ser estimada por escalas a través de las siguientes metodologías:

AUTOR	CLASE	SIGNIFICADO
Tansley & Chipp	1. Rara 2. Ocasional 3. Frecuente 4. Abundante 5. Muy abundante	
Braun - Blanquet	1. Muy escasas 2. Escasas 3. Poco numerosas 4. Numerosas 5. Muy numerosas	
Hanson	1. Escasa 2. Poco frecuente 3. Frecuente 4. Abundante 5. Muy abundante	1-4 plantas /m <sup>2</sup> 5-14 plantas/ m <sup>2</sup> 15-29 plantas/ m <sup>2</sup> 30-99 plantas/ m <sup>2</sup> 100 o más plantas/ m <sup>2</sup>
Acocks	1. Muy rara 2. Rara 3. Ocasional 4. Poco frecuente 5. Frecuente 6. Común 7. Abundante 8. Muy abundante 9. Extremadamente Abundante	1 individuo/Ha 100 m espaciamento 2 70 m 40 17 m 450 5 m 1 individuo/ m <sup>2</sup> 1 m 7 30 cm 40 15 cm 150 7,5 cm 1300 2,5 cm
Barkman, Doing y Segal	1 2 3 4 5 6 7 8 9 10	10 individuos/Ha 1 individuo/100 m <sup>2</sup> 1 individuo/10 m <sup>2</sup> 1 individuo/1 m <sup>2</sup> 1 individuo/10 dm <sup>2</sup> 1 individuo/1 dm <sup>2</sup> 1 individuo/10 cm <sup>2</sup> 1 individuo/1 c m <sup>2</sup> 1 individuo/10 mm <sup>2</sup> 1 individuo/1 mm <sup>2</sup>

**2. Cobertura:** Porcentaje del terreno ocupado por la proyección vertical de una planta, puede ser determinada a través de mapas de las unidades de muestreo, fotografías verticales y la medición se puede realizar por planímetro, grilla de puntos o cuadrículas. Aunque la cobertura se puede calcular existen diversas escalas como son:

AUTOR	CLASE	SIGNIFICADO	
Kylin y Samuelson	1	<8.3%	<1/12
	2	8.3-16.7%	1/12-1/6
	3	16.7-33.3%	1/6-1/3
	4	33.3-66.7%	1/3-2/3
	5	66.7-100%	2/3-1
Braun - Blanquet	1	<10%	<1/10
	2	10-25%	1/10-1/4
	3	25-50%	1/4-1/2
	4	50-75%	1/2-3/4
	5	75-100%	3/4-1
Hult - Sernander	1	<5%	<1/20
	2	5-10%	1/20-1/10
	3	10-20%	1/10-1/5
	4	20-50%	1/5-1/2
	5	50-100%	1/2-1
Pignati	( + )	<1%	
	1	1-20%	
	2	20-40%	
	3	40-60%	
	4	60-80%	
5	80-100%		

**3. Area Basal:** El área o cobertura basal es la superficie de una sección transversal del tallo de la planta evaluada a determinada altura del suelo. Es una medida muy utilizada a nivel forestal y está expresada en  $\frac{m^2}{Ha}$  y la altura de referencia utilizada es de 1,3 m (Graf y Sayagués, 2000) y se conoce como diámetro a la altura del pecho o simplemente DAP. La suma de las secciones de tronco pertenecientes a la misma especie es el área basal correspondiente a dicha especie en un área determinada. El área basal puede ser calculado a partir del DAP promedio y la densidad de población de los árboles (árboles/Ha) aplicando la siguiente fórmula:

$$AB = \frac{\pi}{4} * \overline{DAP}^2 * Densidad$$

Para plantas herbáceas y arbustos ramificados desde la base, la medición del área basal se lleva a cabo en la base del suelo.

**4. Biomasa:** Esta es el peso seco (o fresco, eventualmente) del material vegetal por unidad de área es una variable de amplia aplicación y practicidad. Cuando ésta es registrada a través del tiempo se puede estimar la productividad primaria neta de los ecosistemas. Generalmente, la biomasa es calculada a la parte aérea de las plantas, excluyéndose la parte radicular, es también frecuente fraccionar la biomasa de parte aérea en sus órganos como flores, tallos, follaje.

No obstante, su recolección para posterior secado y el hecho de ser un muestreo destructivo son factores que limitan su aplicación. Aunque, es posible estimar biomasa a través de ecuaciones en función de medidas como diámetro del tronco, ramas y copa y la altura (Graf y Sayagués, 2000), tales ecuaciones son conocidas como "ecuaciones alométricas".

**5. Frecuencia:** Como su nombre lo indica, es la proporción de unidades muestrales donde aparece una especie en relación al número total de unidades evaluadas.

La frecuencia ayuda a determinar la homogeneidad o heterogeneidad de la vegetación y es fundamental para análisis de patrón espacial y uniformidad.

Posee la ventaja que es de fácil y rápida aplicación, pero los datos dependen de la calidad del muestreo y de las características de sus unidades.

**6. Variables Combinadas:** Cuando se toman medidas que pueden ser complementarias, tomadas a la vegetación, son consideradas como semicuantitativas, ya que los valores asignados a cada especie resultan

en una escala ordinal o secuencia de estimaciones conocida como "ranking" (Graf y Sayagués, 2000).

Estas variables pueden ser definidas en función de la densidad y cobertura, abundancia, tipo de distribución espacial. Al respecto existen escalas establecidas para el registro de especies en una comunidad:

AUTOR	CLASE	SIGNIFICADO
Braun - Blanquet	R	individuos solitarios, esporádicos, cobertura insignificante
	( + )	individuos escasos, cobertura <1%
	1	1-10% de cobertura
	2	10-25% de cobertura
	3	25-50% de cobertura
	4	50-75% de cobertura
Tansley	5	75-100% de cobertura
	Dom (dominante)	una planta cubre más que todas las demás juntas
	Codom (codominante)	un pequeño grupo de plantas cubren juntas más de la mitad del área
	ab (abundante)	cobertura menor al 50% pero alta o plantas muy Numerosas
	fr (frecuente)	relativamente baja cobertura pero más bien numerosa
	sp (esparcida)	baja cobertura, escaso número
r (rara)	sólo uno o pocos individuos	
loc (localmente)	puede combinarse con cualquier otra categoría excepto "r"	

## 2. PARÁMETROS Y APLICACIONES DEL MUESTREO PARA SISTEMAS ECOLÓGICOS

### **Antecedentes.**

Cuando se va a llevar a cabo una investigación de carácter ambiental nos enfrentamos al reto de tomar información que sea confiable y que explique el comportamiento de la población de aquello que estamos evaluando; como afirman Mendenhall y Sincich (1997) que, cuando las poblaciones de individuos son grandes ( $>100$ ) es difícil tomar los datos adscritos a tales poblaciones y es necesario tomar un grupo representativo conocido como muestra, la cual permite la comprensión de dicha población; Burrough (1991) asevera que el objetivo del muestreo es el de hacer inferencias sobre la población de interés, basado en la información contenida en la muestra. Ampliando lo anterior se puede indicar que la finalidad que tiene toda muestra es la revelar información sobre la población que representa, de tal forma que se puedan hacer recomendaciones adecuadas con un determinado nivel de confianza. El muestreo de vegetación es la técnica que permite obtener información cualitativa o cuantitativa de la cobertura vegetal de un área determinada, sin necesidad de analizarla o recorrerla en su totalidad (Graf y Sayagués, 2000).

No obstante, se generan diferentes cuestionamientos, tales como:

- ¿Qué muestreamos?
- ¿Cómo muestreamos?
- ¿Porqué muestreamos?
- ¿Cuántas veces muestreamos?

Para entender el muestreo es necesario tener en cuenta algunas consideraciones previas, ya que él está es función de varios aspectos comenzando por el objetivo general de la investigación planteada y el grado de exactitud y precisión (trazabilidad) que se quiere alcanzar con la información obtenida, por tal motivo, el tipo de estudio tiene gran influencia en el muestreo a utilizar, dependiendo, según Graf y Sayagués (2000) de aspectos a tener en cuenta como son:

- Tipo de vegetación y área estudiada.
- Conocimiento previo que se disponga de la flora y vegetación de la zona.
- Concepción o modelo teórico acerca de la naturaleza de las comunidades que adopte el investigador.
- Recursos humanos, económicos, tecnológicos y de tiempo disponibles para desarrollar el estudio.
- Relación entre costos y beneficios obtenidos.
- Necesidad de adecuarse a protocolos de trabajo precedentemente empleados a los efectos de poder realizar comparaciones.

Es de resaltar que la confiabilidad de los resultados depende en gran parte de la representatividad de la muestra (Ovalles, 1992), que se convierte en un aspecto crítico ya que ella actúa como un sustituto y se convierte en la fuente básica de información sobre la cual se basan todos los posteriores análisis e interpretaciones (Size, 1987).

Los métodos de muestreo más aplicados en investigaciones para análisis ambiental son los siguientes:

### ***2.1. Muestreo dirigido, no aleatorio.***

Este método se basa en la selección de lugares típicos. Estos lugares se eligen en base a estudios previos o bien el grado de conocimiento que se tenga sobre lo que se está investigando en la zona. Los sitios seleccionados se consideran representativos de la condición que se quiere caracterizar y generalmente posee un bajo número de observaciones.

Este método tiene como ventaja que es más preciso que otros métodos cuando el tamaño de la muestra es pequeño. Sin embargo, tiene como desventajas que no se puede evaluar la confiabilidad y existe el riesgo de sesgo. Este método solo se recomienda cuando el tiempo y el presupuesto son limitados y si se tiene un muy buen conocimiento del área que se quiere muestrear.

## 2.2. Muestreo aleatorio simple.

Este muestreo se aplica cuando la población no es discriminada por ninguna característica que pueda influir la investigación que se va a llevar a cabo y gira en torno, como su nombre lo indica, hacia la aleatorización de los individuos que se evalúan. Este muestreo se aplica cuando una muestra de tamaño  $n$  es tomada de una población de tamaño  $N$  de tal forma que cada muestra posible de tamaño  $n$  tiene la misma probabilidad de ser seleccionada.

Este muestreo se realiza de la siguiente manera:

- Para variables discretas:

$$n = \frac{(Z_{\alpha/2})^2 \cdot P \cdot Q}{E^2} = \text{aproxima al entero más cercano}$$

Donde:

$n$ : Tamaño de muestra estimado

$Z_{\alpha/2}$ : Distribución normal standard a un nivel medio de significancia.

Confiabilidad (%)	Significancia (%)	Distribución Normal Standard / 2
90	10	1.64
95	5	1.96
99	1	2.57

$P$ : Casos de interés para el investigador, Este valor es dado en forma de proporción, aunque, si no hay reportes al respecto, dicho valor se asume igual a 0.5 (50% de Probabilidad que se presente el caso de interés).

$Q$ : Casos de no interés para el investigador, es el complemento de  $P$ , dado que

$$Q = 1 - P$$

*E* : Error máximo permisible, el cual no debe ser confundido con el nivel de significancia. Este error está basado en características como dinero disponible para alcanzar la precisión del muestreo, calidad del análisis y analista-s estadístico-s, disponibilidad de tiempo, entre otras y tiene un valor máximo del 10%. En condiciones normales de investigación, este valor es del 5%.

Después de calcular la fórmula, se realiza la prueba de corrección por finitud dividiendo el tamaño de muestra por el tamaño de la población y:

Si  $\frac{n}{N} < 0.05$  Trabaja con la muestra calculada.

Si  $\frac{n}{N} \geq 0.05$  Se debe transformar el tamaño de la muestra

↓

$n_0 = \frac{n}{1 + \frac{n-1}{N}} \approx$  Aproxima al entero más cercano

Finalizando aquí el proceso.

*Ejemplo:*

Se tiene un total (N) de 120 árboles a los cuales se les investigará su potencial como hospedero de orquídeas, se debe calcular la muestra a tomar con una confiabilidad del 95% y se sabe además por reportes de otras investigaciones que 4 de cada 10 árboles posee dicha característica, hallar la muestra a tomar con un error máximo permisible del 10%, dado que existe limitación en tiempo.

El procedimiento es:

Confiabilidad=95%, entonces significancia=5%, es decir que, distribución normal standard a nivel medio de significancia es igual a 1.96.

P=4 de cada 10 árboles, entonces  $P = \frac{4}{10} = 0.4$

Q=1-0.4=0.6

Error máximo permisible = 5%, entonces  $E = \frac{10}{100} = 0.1$

Se calcula muestra con base en la fórmula

$$n = \frac{(1.96)^2 * 0.4 * 0.6}{(0.1)^2} = 92.1984 \approx 92 \text{ árboles a muestrear}$$

Luego se calcula la prueba de corrección por finitud:

$$\frac{n}{N} = \frac{92}{120} = 0.77 > 0.05$$

Se debe realizar corrección al tamaño de la muestra, así:

$$n_0 = \frac{92}{1 + \frac{92-1}{120}} = 52.32 \approx 52 \text{ árboles a muestrear}$$

Interpretación: Con una confiabilidad del 95% y un error máximo permisible del 10%, la muestra aleatoria a tomar es de 52 árboles.

- Para variables de índole continuo:

$$n = \frac{(Z_{\alpha/2})^2 * \sigma^2}{H^2} \approx \text{aproxima al entero más cercano}$$

Donde:

n : Tamaño de muestra estimado

$Z_{\alpha/2}$  Distribución normal standard a un nivel medio de significancia.

Confiabilidad (%)	Significancia (%)	Distribución Normal Standard $Z_{\alpha/2}$
90	10	1.64
95	5	1.96
99	1	2.57

$\sigma^2$ : Es la varianza poblacional para la variable de índole continuo bajo estudio.

$H^2$ : Media de anchura, es el intervalo sobre el cual el investigador desea estimar la muestra, se define por criterio del mismo o con base en reportes asociados a la variable que se está evaluando y define el error que el valor promedio deba soportar para que el muestreo posea trazabilidad.

Después de calcular la fórmula, se realiza la prueba de corrección por finitud, de igual manera como para variables discretas, dividiendo el tamaño de muestra por el tamaño de la población y:

Si  $\frac{n}{N} < 0.05$  Trabaja con la muestra calculada.

Si  $\frac{n}{N} \geq 0.05$  Se debe transformar el tamaño de la muestra

↓

$$n_0 = \frac{n}{1 + \frac{n-1}{N}} \approx \text{Aproxima al entero más cercano}$$

Finalizando aquí el proceso.

Ejemplo: Se desea estimar la acumulación de biomasa seca de hoja

para cierta planta que es nativa, se sabe por diferentes reportes que la biomasa seca de hoja de la planta en cuestión posee una desviación standard poblacional de 5 gramos, hallar la muestra a tomar de un total de 210 plantas, para calcular el valor promedio de la biomasa, que sea correcto en más o menos 3 gramos.

El procedimiento es:

- Como no se determinó el nivel de confiabilidad, este debe ser asumido con un valor del 95%, es decir, que la significancia es igual a 5%, por ello  $Z_{\alpha/2} = 1.96$
- Aunque no fue dada la varianza, ésta es la desviación standard al cuadrado:

$$\sigma = 5 \rightarrow \sigma^2 = 5^2 = 25$$

- Exactitud de la prueba,  $H=3$

Se calcula el tamaño de la muestra

$$n = \frac{(1.96)^2 * 25}{(3)^2} = 10.67 \approx 11 \text{ plantas a evaluar}$$

Se realiza la prueba de corrección por finitud:

$$\frac{n}{N} = \frac{11}{210} = 0.0524 > 0.05 \text{ Se debe transformar la muestra.}$$

$$n_0 = \frac{11}{1 + \frac{11-1}{210}} = 10.5 \approx 11$$

Interpretación: La muestra a tomar es de 11 plantas, con una confiabilidad del 95% y una exactitud de  $\pm 3$  gramos para la variable biomasa seca de hoja.

### 2.3. Muestreo aleatorio estratificado.

Este muestreo se aplica cuando la población posee características de importancia, debiéndose estratificarla, es decir, dividir por subpoblaciones, aunque es de resaltar que, al interior de cada subpoblación debe existir homogeneidad entre individuos.

Este muestreo se lleva a cabo dividiendo a los elementos de la población en estratos y seleccionando una muestra aleatoria simple en cada estrato.

Este método de muestreo se recomienda cuando la población no es uniforme, pudiéndose establecer claramente grupos de elementos que constituyen los estratos. Cada estrato debería ser lo suficientemente uniforme de tal manera que la variabilidad entre estratos sea mayor que la variabilidad dentro de los estratos.

Este muestreo se calcula como un muestreo aleatorio simple, pero posee al final un paso adicional que es la distribución equitativa del tamaño de la muestra por subpoblación a través de reglas de tres simples:  $N \rightarrow n$

$$N_x \rightarrow x$$

y se calcula  $x = \frac{N_x * n}{N} \approx$  Aproxima.

*Ejemplo:*

En un bosque no intervenido hay 1235 plantas en total, que están distribuidas de la siguiente manera:

Especie A: 250 plantas

Especie B: 340 plantas

Especie C: 175 plantas

Especie D: 470 plantas

Hallar la muestra tomar con un error máximo permisible del 10% dado que el estudio de índole exploratorio, basado en la morfología de dichas especies.

Dado que la confiabilidad no fue dada, se asume del 95%, y  $Z^{\alpha/2} = 1.96$

P (casos de interés) desconocido, se asume igual a 0,5 (50% de probabilidad), por ende, Q es igual a 0,5 por ser complemento de P. Error máximo permisible igual a 0,1.

El procedimiento es:

$$n = \frac{(1.96)^2 * 0.5 * 0.5}{(0.1)^2} = 96.04 \approx 96 \text{ plantas a tomar}$$

Se realiza la prueba de corrección por finitud:

$$\frac{n}{N} = \frac{96}{1235} = 0.078 > 0.05$$

Debe transformarse el tamaño de la muestra así.

$$n_0 = \frac{96}{1 + \frac{96 - 1}{1235}} = 89.14 \approx 89 \text{ plantas a tomar para el muestreo.}$$

Finalmente se realiza la estratificación por especie:

Especie A.

$$n_{sppA} = \frac{89 * 250}{1235} = 18.02 \approx 18$$

Especie B.

$$n_{sppB} = \frac{89 * 340}{1235} = 24.5 \approx 25$$

Especie C.

$$n_{sppC} = \frac{89 * 175}{1235} = 12.61 \approx 13$$

Especie D.

$$n_{sppD} = 89 - 18 - 25 - 34 = 33$$

Interpretación: La muestra a tomar con una confiabilidad del 95% y un error máximo permisible del 10% es de 89 plantas, distribuidas así:

Especie A: 18 plantas

Especie B: 25 plantas

Especie C: 13 plantas

Especie D: 33 plantas

#### **2.4. Muestreo sistemático o secuencial.**

Este tipo de muestreo se aplica cuando se toma a los individuos aleatorios en secuencia, maneja el criterio de los muestreos aleatorio simple y estratificado,

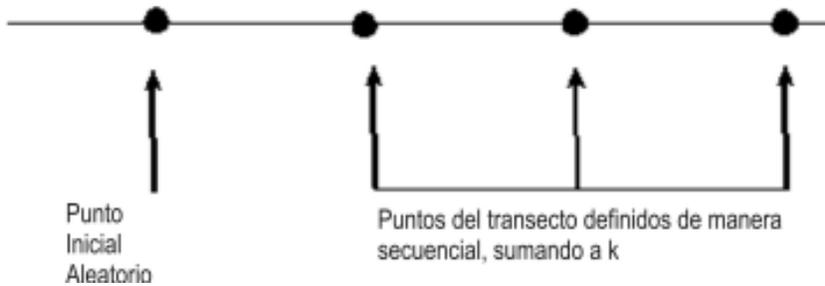
Este método de muestreo consiste en tomar las muestras a distancias fijas conocidas como transectos y posee variantes, tales como:

- *Muestreo sistemático unidimensional:*

La muestra se obtiene mediante la selección aleatoria de un elemento de los primeros k elementos en una línea, donde se define primero el tamaño de la muestra a través del muestreo aleatorio simple, seguidamente se invierte la proporción de la corrección por finitud, quedando de la siguiente manera:

$$\frac{N}{n} \approx \text{aproxima al entero más cercano, y es el valor que define al punto}$$

de partida que equivale a la primer unidad de muestreo, por medio de un proceso estocástico de números aleatorios y con base en el tamaño de k se determina la secuencia a seguir para definir los demás puntos de la siguiente manera: número aleatorio, número aleatorio+k, resultado anterior+k, y así sucesivamente se repite la operación hasta definir las muestras a tomar, como su nombre lo indica, de manera secuencial.



*Ejemplo:*

Se determinó que el tamaño de muestra para un transecto de longitud de 3500 metros es de 59, sabiendo que el total de unidades de muestreo posibles es de 350 (de 10 metros cada una), determine la secuencia:

$$k = \frac{N}{n} = \frac{150}{59} = 2.54$$

Indica que se debe escoger en forma aleatoria un número entre 1 y 2,54 para establecer el primer sitio.

Suponiendo que se tomó el número 2 (A los 20 metros de inicio del transecto) y sigue:

Unidades: 2,  $2 + 2.54 = 4.54$ ,  $4.54 + 2.54 = 7.08$ ,  $7.08 + 2.54 = 9.62$ , 12, 16, 14. 7, 17.24, 19.78, 22.32, 24.86, 27.4, 29.94, 32.48, ... , 149.32

Sin embargo, no pueden ser tomadas las muestras (en este caso los transectos) con decimales; por ello deben aproximarse todos los valores al entero más cercano y se define así para el ejemplo: 2, 5, 7, 10, 12, 15, 17, 20, 22, 25, 27, 30, 32, ..., 150.

- *Muestreo sistemático bidimensional:*

Los sitios de muestreo son definidos a intervalos regulares en una rejilla, cuadrícula o parrilla (llamada en inglés "grid") y las observaciones se toman en las intersecciones de la rejilla o en el centro de cada celda que se genera. Es de resaltar que, dicha rejilla está conformada por celdas de igual tamaño y forma.

Para desarrollar esta técnica, el primer punto se selecciona aleatoriamente (como se explicó en el caso anterior) y las demás muestras se toman en distancias fijas en ambas direcciones.

UM1	UM2	UM3	UM4
UM5	UM6	UM7	UM8
UM9	UM10	UM11	UM12
UM13	UM14	UM15	UM16

En la gráfica se aprecian la rejilla y 16 posibles unidades de muestreo que pueden ser determinados.

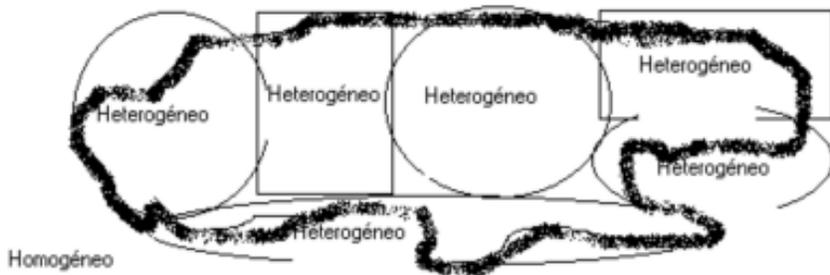
Es de resaltar que el muestreo secuencial ha cobrado suma importancia en las investigaciones en ecología, dado que es muy posible que la población bajo estudio sea heterogénea o sensible a factores ambientales, y es por lo anteriormente expuesto que, se recomienda éste dado que se acopla a la ubicación en mapas para el caso de análisis de índole exploratorio.

### **2.5. Muestreo por conglomerados.**

Este tipo de muestreo se aplica cuando se conforman grupos, que en general, son homogéneos entre ellos, pero al interior de cada uno de ellos es heterogéneo.

Este muestreo se basa en un aleatorio simple en el cual cada unidad de muestreo es un grupo de elementos bajo evaluación y se aplica cuando se realiza la investigación en áreas grandes y existe limitación de recursos y tiempo para abarcarlas.

Este tipo de muestreo requiere que la variabilidad dentro de los conglomerados sea mayor que la variabilidad entre conglomerados.



## **2.6. Muestreo Anidado.**

Este tipo de muestreo está basado en el diseño experimental jerárquico o anidado, que se basa en el criterio de submuestreo y es útil cuando la variable que se está evaluando es susceptible de variabilidad en función del tiempo o del espacio, lo cual indica desconocimiento acerca de la población bajo estudio.

El muestreo anidado está basado en la repetición del muestreo a diferentes intervalos para detectar la variabilidad en cada uno de ellos, lo cual es corroborado por Montgomery (1991), cuando indica que un diseño anidado puede usarse para identificar las fuentes principales de variabilidad en la respuesta. Lo cual determina el cambio de la misma en los diferentes intervalos de muestreo y de esta manera definir el comportamiento real y la dinámica de la variable bajo estudio.

Este método posee la ventaja de ser utilizado como muestreo preliminar, permitiendo estimar la distancia de muestreo más eficiente, además del área mínima más adecuada, que es definida por Graf y Sayagués (2000) como la menor superficie de terreno que contiene la casi totalidad de las especies de una determinada comunidad.

No obstante, debe determinarse el tamaño de las unidades con base en el siguiente modelo:

			7		
		5			
				6	
	3		4		
1	2				

El diseño es de la forma (Montgomery,1991):

$$y_{ijkl} = \mu + \tau_i + \beta_{j(i)} + \gamma_{k(ij)} + \varepsilon_{(ijkl)} \begin{cases} i = 1,2,3,\dots,a \\ j = 1,2,3,\dots,b \\ k = 1,2,3,\dots,c \\ l = 1,2,3,\dots,n \end{cases}$$

Donde:

$\mu$ : Efecto promedio de la variable

$\tau_i = A$ : Efecto del factor A.

$\beta_{j(i)} = B(A)$ : Efecto del factor B dentro del factor A.

$\gamma_{k(ij)} = C(B)$ : Efecto del factor C dentro del factor B.

$\varepsilon_{(ijkl)} = ErrorExperimental = D(C)$ : Efecto del factor D dentro del factor C, es lo mismo que el error experimental.

Un número de niveles adecuado es cuatro, debido a que es la cifra mínima que permite detectar cambios consistentes en la pendiente del variograma que se genere (Ovalles, 1992), por lo tanto, en estudios de índole ambiental se maneja siempre hasta el factor D, aunque por

cuestiones de operatividad en la toma de datos y como regla arbitraria se ha utilizado el sistema 7x2x2x2, el cual indica 7 áreas de factor A, 2 áreas de factor B dentro de cada área de factor A, 2 áreas de factor C, dentro de cada área de factor B obtenida y 2 áreas de factor D dentro de cada área de factor C obtenido, generándose 56 unidades de muestreo. Aunque es posible también manejar el sistema 2x2x2x2, el cual es más operativo y económico a pesar de tener la desventaja de correr mayor riesgo de imprecisión en el análisis.

El análisis de varianza (ANAVA) para el diseño anidado es:

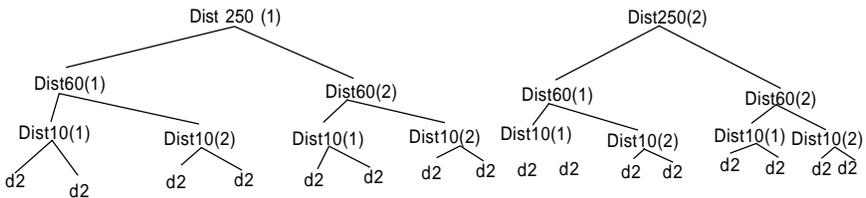
### ANAVA

FV	GL	SC	CM	$t_{Calculado}$	$F_{Tabla, 5\%, 1\%$
A	$a - 1$	$\sum_j \frac{y_{1..}^2}{bcn} - \frac{(\sum y_{ij})^2}{n}$	$\frac{SC_A}{GL_A}$	$\frac{CM_A}{CM_{B(A)}}$	$(GL_A, GL_{B(A)})$
B(Dentro de A)	$a(b-1)$	$\sum_j \sum_j \frac{y_{j..}^2}{cn} - \sum_j \frac{y_{1..}^2}{bcn}$	$\frac{SC_{B(A)}}{GL_{B(A)}}$	$\frac{CM_{B(A)}}{CM_{C(B)}}$	$(GL_{B(A)}, GL_{C(B)})$
C(dentro de B)	$ab(c-1)$	$\sum_j \sum_j \sum_k \frac{y_{ijk}^2}{n} - \sum_j \sum_j \frac{y_{j..}^2}{cn}$	$\frac{SC_{C(B)}}{GL_{C(B)}}$	$\frac{CM_{C(B)}}{CM_{ErrorExp}}$	$(GL_{C(B)}, GL_{Error})$
Error Exp.	$abc(n-1)$	$\sum_j \sum_j \sum_k \sum_k \frac{y_{ijkl}^2}{n} - \sum_j \sum_j \sum_k \frac{y_{ijk}^2}{cn}$	$\frac{SC_{ErrorExp}}{GL_{ErrorExp}}$		
Total	$n-1$	$\sum y_{ij}^2 - \frac{(\sum y_{ij})^2}{n}$			

Como complemento al análisis de varianza se construye el gráfico de las varianzas acumuladas contra las distancias expresadas logarítmicamente y conocido como variograma, el cual es de amplia aplicación en el estudio de la geoestadística, cuyo criterio es que la distancia donde se observe un cambio marcado en la pendiente de la curva, constituye la distancia que mejor resuelve el patrón de variación de la variable (Ovalles, 1991).

**Ejemplo:**

En una investigación se desea determinar el comportamiento del peso en gramos de biomasa seca de follaje de cierto arbusto en función del espacio (Al interior de una sucesión vegetal). Para el estudio se establecieron arbitrariamente 2 distancias de 250 metros (factor A), de esas 2, se tomaron 2 distancias de 60 metros (factor B), a su vez, del factor B se tomaron 2 distancias de 10 metros (factor C) y por último se tomaron 2 distancias, extraídas del factor C (factor D) y se tomaron 2 lecturas (repeticiones) por unidad de muestreo, dado que  $2 \times 2 \times 2 \times 2 = 16$  unidades de muestreo.



Cuyos resultados son (discriminado por cada d2):

Dist250 (1)							
Dist 60 (1)				Dist 60 (2)			
Dist 10 (1)		Dist 10 (2)		Dist 10 (1)		Dist 10 (2)	
d2	d2	d2	d2	d2	d2	d2	d2
5.1	5.6	6.1	5.8	5.4	5.7	5.9	6
5.3	5.7	6.3	5.9	5.5	5.8	5.9	6.1

Dist 250 (2)							
Dist 60 (1)				Dist 60 (2)			
Dist 10 (1)		Dist 10 (2)		Dist 10 (1)		Dist 10 (2)	
d2	d2	d2	d2	d2	d2	d2	d2
6.2	6	5.5	5.7	5.5	5.3	5.4	5.6
6.1	6.3	5.4	5.3	5.4	5.5	5.2	5.1

Pasos a seguir:

1. Se calcula la suma de cuadrado total:

$$SC_{Total} = 5.1^2 + 5.3^2 + \dots + 5.1^2 - \frac{(5.1 + 5.3 + \dots + 5.1)^2}{32} = 1034.32 - 1030.58 = 3.74$$

2. Se calcula la suma de cuadrado del factor A (distancia de 250 m):

$$SC_A = \frac{(5.1 + 5.3 + \dots + 6.1)^2 + (6.2 + 6.1 + \dots + 5.1)^2}{16} - 1030.58 = 1030.79 - 1030.58 = 0.21$$

3. Se calcula la suma de cuadrado de B(A) (Distancia de 60 dentro de 250 m):

$$SC_{B(A)} = \frac{(5.1 + \dots + 5.9)^2 + (5.4 + \dots + 6.1)^2 + (6.2 + \dots + 5.3)^2 + (5.5 + \dots + 5.1)^2}{8} - 1030.79 \rightarrow$$
$$SC_{B(A)} = 1031.57 - 1030.79 = 0.78$$

4. Se calcula la suma de cuadrado de C(B) (Distancia de 10 dentro de 60 m):

$$SC_{C(B)} = \frac{(5.1 + \dots + 5.7)^2 + (6.1 + \dots + 5.9)^2 + \dots + (5.4 + \dots + 5.1)^2}{8} - 1031.57 \rightarrow$$
$$SC_{C(B)} = 1033.51 - 1031.57 = 1.94$$

5. Se calcula la suma de cuadrado de D(C) (Distancia de 2 dentro de 10 m):

$$SC_{D(C)} = \frac{(5.1 + 5.3)^2 + (5.6 + 5.7)^2 + \dots + (5.6 + \dots + 5.1)^2}{8} - 1033.51 \rightarrow$$
$$SC_{D(C)} = 1033.95 - 1033.51 = 0.82$$

6. Se construye el ANAVA así:

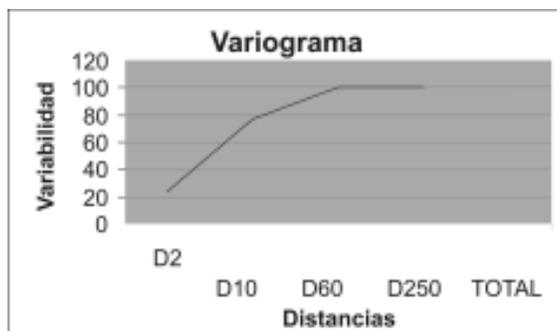
## ANAVA

FV	GL	SC	CM	$f_{\text{Calculado}}$	$F_{\text{Tabla, 5\%, 1\%}}$
A	2-1=1	0.21	$\frac{0.21}{1} = 0.21$	$\frac{0.21}{0.39} = 0.54$	(1,2) = 18.51 y 98.5
B(Dentro de A)	2(2-1)=2	0.78	$\frac{0.78}{2} = 0.39$	$\frac{0.39}{0.485} = 0.8$	(2,4) = 6.94 y 18
C(dentro de B)	2*2(2-1)=4	1.94	$\frac{1.94}{4} = 0.485$	$\frac{0.485}{0.0342} = 14.18$ (**)	(4,24) = 2.78 y 4.22
D(dentro de C) <sub>1</sub>	31-1-2-4=24	0.21-0.78-1.94=0.82	$\frac{0.82}{24} = 0.0342$		
Total	32-1=31	3.74			

### 1. Equivalente a Error Experimental

Interpretación: Se aprecia que existe diferencia significativa para C(D), lo cual indica que para la distancia 10(60) m se encuentra la mayor variabilidad, indica que existe gran incremento entre 10 y 60 m. Muestrear a distancias mayores de 60 m resulta en una información demasiado variable. Muestrear a distancias menores de 10 m resulta en un muestreo innecesariamente intenso.

Indica que se puede muestrear entre 10 y 60 m, la biomasa seca de follaje del arbusto evaluado.



### **3. ANÁLISIS DE PATRÓN ESPACIAL.**

#### **Antecedentes.**

El análisis de patrón espacial de plantas y animales es una característica importante de comunidades ecológicas, la cual es una de las primeras observaciones que se deben llevar a cabo para comprender así diferentes organismos vivos.

Se reconocen tres tipos de patrones en las comunidades: aleatorio, agregado y uniforme, permitiendo el planteamiento de hipótesis que permiten resaltar los posibles factores causales para que estos se hayan presentado. Una vez establecido el patrón, la generación de la hipótesis concierne a la estructura ecológica de las comunidades.

El patrón aleatorio en una población de organismos implica homogeneidad ambiental y/o patrones ambientales no selectivos. Por otro lado, los patrones agregado y uniforme indican que existen algunas restricciones que determinan el comportamiento de la comunidad ecológica. La agregación sugiere que los individuos están más reunidos en las partes más favorables del hábitat, esto puede ser debido a comportamiento gregario de la especie, heterogeneidad ambiental, modo reproductivo, entre otros. El patrón de distribución espacial uniforme resulta de interacciones negativas entre individuos, tales como competencia por comida o espacio.

Según Hutchinson (1953), dentro de los posibles factores causales de distribución espacial de organismos están:

- Factores vectoriales resultantes de la acción de las condiciones ambientales circundantes.
- Factores de comportamiento territorial o de índole social.
- Factores reproductivos de la especie bajo estudio.
- Factores coactivos resultantes de las interacciones intraespecíficas (competencia).
- Factores estocásticos resultantes en la variación aleatoria en un factor precedente.

Con base en lo expuesto anteriormente se puede afirmar que los tipos de distribución espacial están sujetos a factores intrínsecos y extrínsecos con respecto a la especie.

### 3.1. Métodos de distribución.

Para analizar más a fondo los tipos de distribución espacial, se debe tener en cuenta un criterio el cual es la dispersión aleatoria, que indica que cada sitio (Unidad de muestreo) posee igual oportunidad de hospedar un individuo y este no influye sobre otros que podrían estar presentes.

La determinación de los patrones espaciales mencionados con antelación posee una prueba inicial simple basada en la media aritmética y la varianza para el número de individuos por unidad experimental e indicando el tipo de distribución, de la siguiente manera:

1. Patrón aleatorio: media aritmética = varianza      Distribución de Poisson.
2. Patrón agregado: media aritmética < varianza      Distribución Binomial negativa.
3. Patrón uniforme: media aritmética > varianza      Distribución Binomial positiva.

El primer cuestionamiento para detectar un patrón en una comunidad ecológica es probar la hipótesis que la distribución de los individuos por unidad de muestreo es de índole aleatorio, si esta es rechazada, la distribución podría estar orientada hacia los patrones agregado o uniforme.

Procedimientos para determinar el análisis de patrón espacial:

\*Distribución de probabilidad de Poisson.

Para el modelo probabilístico de Poisson en función del número de individuos por unidad de muestreo existen las siguientes condiciones (Ludwing y Reynolds, 1988):

1. Cada unidad de muestreo natural posee una probabilidad igual de hospedaje de un individuo.
2. La ocurrencia de un individuo en una unidad de muestreo no está influenciada por ocupación por otra.
3. Cada unidad de muestreo es igualmente disponible.
4. El número de individuos por unidad de muestreo es relativamente bajo al máximo posible que pudo ocurrir en tal unidad de muestreo.

Para el cálculo de las distribuciones, se estima primero el número de individuos promedio por unidad de muestreo y con base en los siguientes pasos:

1. Establecer la hipótesis: Número de individuos que determinen una distribución de Poisson.
2. Se toman los datos a mínimo 30 unidades de muestreo para que el análisis tenga solidez y se minimice el riesgo de sesgo o error.
3. Se calcula el modelo probabilístico de Poisson así:

$$P(x) = \frac{\mu^x * e^{-\mu}}{x!}$$

Aplicando sumatoria para cuantas observaciones o individuos x haya, dado que

$P(x)=1, 2, 3, \dots, x.$

4. Se calculan las frecuencias esperadas de Poisson  $E(x)$ , donde cada probabilidad calculada de Poisson se multiplica por el número total  $n$  de unidades de muestreo en la muestra bajo estudio, aplicando la siguiente fórmula:

$$E(x) = n * P(x)$$

Cada una de las ecuaciones representa una clase de frecuencia, obteniéndose un total  $q$  de  $x+1$  clases de frecuencia de individuos esperados.

Como regla general, los valores esperados de  $x$  ( $P(x)$ ), deben ser siempre menor que 1, dado que la sumatoria de los valores de las probabilidades de  $x$  debe ser igual a 1 (100% de probabilidad).

5. Prueba de bondad de ajuste para la prueba Chi-cuadrado ( $X^2$ ): Determina que tan acertadas son las frecuencias observadas  $F(x)$  con respecto a las frecuencias esperadas  $E(x)$ . La distribución  $X^2$  se calcula así:

$$X^2 = \sum_{x=0}^q \frac{(f(x) - E(x))^2}{E(x)}$$

El resultado del cálculo de esta fórmula se compara con el valor de  $X^2$  de tabla con  $q - 2$  grados de libertad (recordar que  $q$  es igual a  $x + 1$ ) y bajo cierto nivel de significancia como es el 5% o 1%, según el criterio del investigador, aunque se aplica con mayor frecuencia el primero.

Si  $X^2$  calculado  $>$   $X^2$  tabla bajo el nivel de significancia aplicado, se concluye que el patrón no es de índole aleatorio (no sigue la distribución de Poisson), rechazándose la hipótesis planteada inicialmente.

Si  $X^2$  calculado  $<$   $X^2$  tabla bajo el nivel de significancia aplicado, se concluye que el patrón es aleatorio, aceptándose la hipótesis.

\*Distribución de probabilidad Binomial negativa.

Este modelo probabilístico es el más usado para determinar agregación de poblaciones.

Cuando las dos condiciones establecidas para el modelo de Poisson no son satisfechas, usualmente sugiere que pueda existir un patrón agregado.

El modelo probabilístico binomial negativo posee dos parámetros:

1.  $\mu$ , el número promedio de individuos por unidad de muestreo.
2.  $K$ , parámetro relacionado al grado de asociación.

Para esta distribución, se manejan pasos similares a los aplicados a la serie de Poisson:

1. Establecer la hipótesis: Que el número de individuos por unidad de muestreo siga una distribución binomial negativa y además exista un patrón agregado o no aleatorio.
2. Se determina la distribución de frecuencia,  $F(x)$ .
3. Cálculo de la probabilidad binomial negativa,  $P(x)$ : Con base en la probabilidad de encontrar  $x$  individuos en una unidad de muestreo, donde  $x = 1, 2, 3, \dots, n$  y está dado por:

$$P(x) = \left[ \frac{\mu}{\mu + k} \right]^x \left( \frac{(k + x - 1)}{(x!(k - 1))!} \right) \left( 1 + \left( \frac{\mu}{k} \right) \right)^{-k}$$

Donde  $\mu$  es el promedio de la muestra y  $k$  es el grado de agregación y se estima de la siguiente manera:

$$\log_{10}(N / N_0) = \hat{k} \log_{10} \left[ 1 + \left( \frac{\bar{x}}{\hat{k}} \right) \right]$$

Donde  $N$  es el número total de unidades de muestreo y  $N_0$  es el número total de unidades de muestreo con cero individuos.

Para la estimación anterior se compara el valor del lado derecho con el del lado izquierdo hasta que convergen en el mismo valor; no obstante, un buen cálculo de  $k$  estimado para la primera iteración es obtenido así:

$$\hat{k} = \frac{\bar{x}^2}{s^2 - \bar{x}}$$

- Donde  $s^2$  es la varianza estimada para la muestra.
- Cuando  $\mu$  es pequeño (menor que 4), la ecuación anterior es ineficiente para calcular  $k$  estimado.
- Si  $\mu$  es mayor que 4 esta estimación es efectiva si la agregación es extensiva en la población.
- Si la media muestral y  $k$  estimado son mayores que 4, entonces la ecuación anterior es muy adecuada para el cálculo mencionado con antelación.

Después de hallar la media muestral y k estimado se aplica la ecuación así:

$$P(x) = \left[ \frac{x}{x + \hat{k}} \right] \left[ \frac{(\hat{k} + r - 1)}{r!(\hat{k} - 1)} \right] \left[ \frac{x}{\hat{k}} \right]^k$$

4. Se calculan posteriormente las frecuencias binomiales negativas esperadas E(x): El número esperado de unidades de muestreo con x individuos es obtenido multiplicando cada una de las probabilidades binomiales negativas por N (número total de unidades de muestreo en la muestra y q se calcula como se aplicó para el modelo de Poisson.
5. Prueba de bondad de ajuste completa  $X^2$  : Se lleva a cabo aplicando la ecuación

$$X^2 = \sum_{x=0}^q \frac{(f(x) - E(x))^2}{E(x)}$$

y comparando con la tabla de  $X^2$  con  $(q - 2) - 1$  grados de libertad; es decir,  $q - 3$  grados de libertad.

*Ejemplo:*

Se llevó a cabo un estudio para analizar la distribución espacial de larvas de *Spodoptera spp* en un cultivo de maíz de 30 días después de emergencia, el conteo se realizó en sus hojas y se evaluaron 180 unidades de muestreo (180 plantas) y la distribución de frecuencia obtenida fue la siguiente:

No larvas observadas (x)	0	1	2	3	4	5	6	7	8	9	10
No unidades -muestreo (fx)	11 4	25	15	10	6	5	2	1	1	0	1

$$n = \sum_{x=0}^{10} (x * fx) = (0*114)+(1*25)+(2*15)+. . . +(9*0)+(10*1) = 171$$

media aritmética de la muestra:  $x = \frac{n}{N} = \frac{171}{180} = 0.95$

varianza de la muestra:  $s^2 = \left( \sum_{x=0}^{10} (x * fx)^2 - x * n \right) / (n - 1)$

$$\rightarrow s^2 = (681 - (0.95 * 171)) / (171 - 1) = 2.897$$

Paso 1: Se establece la hipótesis (Ho) que las larvas de Spodoptera spp siguen un patrón aleatorio.

Paso 2: Se calcula el modelo probabilístico de Poisson para cada una de las distribuciones (x), como se plantea aquí para x=0,

$$P(0) = \frac{0.95^0 * e^{-0.95}}{0!} = 0.3867$$

Como la aparición de 4 o más larvas contribuye solamente en un 1.6%, es por ello que no es necesario calcular para las demás distribuciones, aglutinándolas en un solo grupo para el cálculo de Poisson.

Paso 3: Posteriormente se calculan los valores esperados para las distribuciones a las cuales se les halló la probabilidad respectiva, así:

$$E(x) = N * P(0) \rightarrow E(0) = 180 * 0.3867 = 69.61$$

Paso 4: Seguido al paso anterior se calcula  $X^2$ , con base en las frecuencias (Fx) y los valores esperados (Ex), el valor obtenido comparado el valor de  $X^2$  de la tabla con q-2 grados de libertad (q = número de cálculos realizados para poisson), si el  $X^2$  calculado es mayor o igual al valor de  $X^2$  de la tabla se rechaza la hipótesis que las larvas siguen una distribución aleatoria.

$$X^2 \text{ calculado} = 121.8$$

$$X^2 \text{ tabla} = 14.067$$

Entonces se rechaza la hipótesis (Ho) que las larvas de *Spodoptera spp* siguen una distribución aleatoria en las plantas de maíz.

Como fue rechazada la hipótesis del patrón aleatorio, entonces se calcula para el patrón agregado con base en el modelo binomial negativo.

Paso 1: Las larvas siguen una distribución agregada

Paso 2: Se calculan las probabilidad binomial negativa

$$\hat{k} = 0.4635$$

Como  $\hat{k}$  y  $\bar{x}$  son menores a 1, entonces se aplica la ecuación :

$$\hat{k} \log_{10} (1 + \bar{x} / \hat{k}) = 0.2245$$

Además N=180 y No=114, entonces  $\log_{10} \left( \frac{180}{114} \right) = 0.1984$

Como  $0.1984 < 0.4635$ , sustituye  $\hat{k}$  por 0.3, entonces

$$\hat{k} \log_{10} (1 + \bar{x} / \hat{k}) = 0.1859$$

Como 0.1859 es más cercano a 0.1984, pero es más bajo, entonces se utiliza un  $\hat{k}$  de 0.3456, entonces

$$\hat{k} \log_{10} (1 + \bar{x} / \hat{k}) = 0.1984$$

Paso 3: Después de determinar  $\hat{k}$  se calcula la probabilidad binomial negativa, sin embargo debe llevarse a cabo para cada x, hasta que haya alguno que acerque al 100%, aglutinando a los demás x.

Cuando  $x=0$

$$P(0) = [1 + (0.95 / 0.3457)]^{0.3457} = 0.6333$$

Paso 4: Se calculan los valores esperados, con base en  $N=180$  y luego se calcula a  $X^2$  y se repite el proceso como en el paso 4 para la distribución de poisson, pero  $X^2$  de la tabla se calcula con base en  $q-3$  grados de libertad (lo cual difiere con respecto a poisson).

Si el valor de  $X^2$  calculado es menor al valor  $X^2$  de la tabla, se acepta la hipótesis de la distribución agregada para las larvas de *Spodoptera spp.*

$$X^2 \text{ calculado} = 1.18$$

$$X^2 \text{ tabla} = 7.82$$

### 3.2 . Métodos de varianza-cuadrante.

Estos métodos son aplicados cuando los individuos se distribuyen en forma continua a través de una comunidad (p.ej. orquídeas en un bosque), es por esto que se debe escoger el tipo de unidad de muestreo de manera adecuada según el criterio del investigador, con base en la forma y tamaño de dicha unidad de muestreo, es de utilidad para el estudio de corredores biológicos.

Estos métodos no afectan las distribuciones espaciales de las especies examinadas y giran en torno a la media aritmética y varianza del número de individuos por unidad de muestreo (UM).

- Método de varianza de bloque (BQV): En este método se calculan las varianzas de los números de individuos en bloques de diferentes tamaños, los cuales son obtenidos por la combinación progresiva de  $N$  cuadros (particiones del campo bajo estudio) de manera prescrita.

Ejemplo:

Se tiene un lote de 100 metros de longitud, determinar los posibles bloques:

Cada 10 metros: (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

Cada 20 metros: (1,2) (3,4) (5,6) (7,8) (9,10)

Cada 50 metros: (1,2,3,4,5) (6,7,8,9,10)

Siempre los bloques deben poseer la misma longitud.

Para calcular BQV se deben llevar a cabo los siguientes pasos:

Paso 1. Se calcula BQV para bloque de tamaño 1 así:

$$VAR(x) = \left(\frac{2}{N}\right) \left\{ \left[ \frac{1}{2} * (x_1 - x_2)^2 \right] + \left[ \frac{1}{2} * (x_3 - x_4)^2 \right] + \dots + \left[ \frac{1}{2} * (x_{n-1} - x_n)^2 \right] \right\}$$

Paso 2. Los datos contenidos en pares adyacentes son combinados para dar 2 bloques cuadráticos que son 2 veces más largos pero es la mitad de la cantidad del anterior, su varianza está dada por:

$$VAR(x) = \left(\frac{4}{N}\right) \left\{ \left[ \frac{1}{4} * (x_1 - x_2 - x_3 - x_4)^2 \right] + \left[ \frac{1}{4} * (x_5 - x_6 - x_7 - x_8)^2 \right] + \dots + \left[ \frac{1}{4} * (x_{n-3} - x_{n-2} - x_{n-1} - x_n)^2 \right] \right\}$$

Paso 3. Se calculan las varianzas para los tamaños de bloques más grandes, con base en el tamaño o longitud del cinturón.

Paso 4. Se agrupan las respectivas varianzas contra sus tamaños de bloques.

Sin embargo, BQV es limitado sólo para pares de 2, es por ello que se desarrolló el método de varianza cuadrática de 2 términos locales (TTLQV), el cual reemplaza el uso de BQV y se muestra en el siguiente ejemplo:

Para los mismos 100 metros:

Bloques de 10 metros: (1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

Bloques de 20 metros: (1,2) (3,4) (5,6) (7,8) (9,10)

Bloques de 30 metros: (1,2,3) (4,5,6) (7,8,9)

Bloques de 40 metros: (1,2,3,4) (5,6,7,8)

Bloques de 50 metros: (1,2,3,4,5) (6,7,8,9,10)

Aquí no es necesaria la misma longitud de los bloques. TTLQV se calcula así:

Paso 1. Tamaño de bloque 1,

$$VAR(x)1 = \left[ \frac{1}{(N-1)} \right] \left\{ \left[ \frac{1}{2} * (x_1 - x_2)^2 \right] + \left[ \frac{1}{2} * (x_2 - x_3)^2 \right] + \dots + \left[ \frac{1}{2} * (x_{N-1} - x_N)^2 \right] \right\}$$

Paso 2. Tamaño de bloque 2,

$$VAR(x)2 = \left[ \frac{1}{(N-3)} \right] \left\{ \left[ \frac{1}{2} * (x_1 + x_2 - x_3 - x_4)^2 \right] + \left[ \frac{1}{2} * (x_2 + x_3 - x_4 - x_5)^2 \right] + \dots + \left[ \frac{1}{2} * (x_{N-3} + x_{N-2} - x_{N-1} - x_N)^2 \right] \right\}$$

El método TTLQV estima las varianzas para todos los tamaños de bloques superiores a un tamaño de bloque superior a N/2 (N/10 es recomendado).

- Método de varianza pareada cuadrática (PQV): Se seleccionan pares de cuadros en un espaciamiento específico o distancia a través de la franja o cinturón para estimar las varianzas. Desde que los cuadros sean de un tamaño fijado, el único componente para la variación es la distancia entre ellos, se calcula así:

Paso1. Se halla la varianza PQV en el espaciamiento 1 por:

$$VAR(x)1 = \left[ \frac{1}{N-1} \right] \left\{ \left[ \frac{1}{2} * (x_1 - x_2)^2 \right] + \left[ \frac{1}{2} * (x_2 - x_3)^2 \right] + \dots + \left[ \frac{1}{2} * (x_{N-1} - x_N)^2 \right] \right\}$$

Paso 2. Varianza en el espaciamiento 2,

$$VAR(x)2 = \left[ \frac{1}{N-2} \right] \left\{ \left[ \frac{1}{2} * (x_1 - x_3)^2 \right] + \left[ \frac{1}{2} * (x_2 - x_4)^2 \right] + \dots + \left[ \frac{1}{2} * (x_{N-2} - x_N)^2 \right] \right\}$$

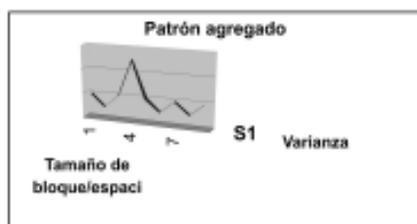
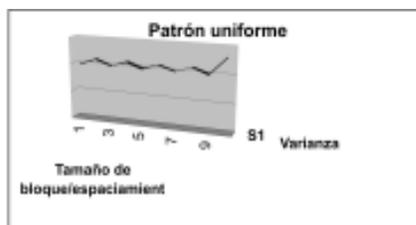
Paso 3. Los cálculos análogos son usados para otras varianzas en espaciamentos más grandes, cuando un espaciamiento seleccionado es menor que el máximo espaciamiento de N-1.

Paso 4. Se contrastan las varianzas obtenidas por el método PQV contra el espaciamiento del cuadro, en el método PQV las varianzas

son calculadas para todos los cuadros pareados posibles en un espaciamiento dado.

La selección aleatoria es sin reemplazamiento, donde las estimaciones de las varianzas de los espaciamientos son independientes.

Interpretación de los planos de varianza:



- En el patrón aleatorio, los individuos están dispersos, como su nombre lo indica, de manera aleatoria a través del área bajo estudio; es decir, las varianzas fluctúan aleatoriamente con los diferentes tamaños de bloques.
- En el patrón uniforme, los individuos están dispersos de manera regular, las varianzas serían bajas y tienden a no fluctuar en los diferentes tamaños de bloques.
- En el patrón agregado, la varianza tiende a un pico, en un tamaño de bloque equivalente al radio del área promedio agregada (área promedio ocupada por cada agregación).

Ejemplo:

El siguiente es el número de plantas encontradas de cierta orquídea en un transecto de 7 cuadrantes contiguos (N=7):

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$
X=	8	6	3	0	4	7	5

Hallar el comportamiento de la orquídea evaluada, con base en los métodos varianza-bloque y varianza pareada cuadrática.

Se resuelve de la siguiente manera:

Método TTLQV:

Varianza en bloque tamaño 1:

$$VAR(x) = \left[ \frac{1}{(7-1)} \right] \left\{ \left[ \frac{1}{2} (8-6)^2 \right] + \left[ \frac{1}{2} (6-3)^2 \right] + \left[ \frac{1}{2} (3-0)^2 \right] + \left[ \frac{1}{2} (0-4)^2 \right] + \left[ \frac{1}{2} (4-7)^2 \right] + \left[ \frac{1}{2} (7-5)^2 \right] \right\} = 4.25$$

Varianza en bloque tamaño 2:

$$VAR(x) \mathcal{B} = \left[ \frac{1}{(7-3)} \right] \left\{ \left[ \frac{1}{4} (8+6-3-0)^2 \right] + \left[ \frac{1}{4} (6+3-0-4)^2 \right] + \left[ \frac{1}{4} (3+0-4-7)^2 \right] + \left[ \frac{1}{4} (0+4-7-5)^2 \right] \right\} = 17.125$$

Varianza en bloque tamaño 3:

$$VAR(x) \mathcal{B} = \left[ \frac{1}{(7-5)} \right] \left\{ \left[ \frac{1}{6} (8+6+3-0-4-7)^2 \right] + \left[ \frac{1}{6} (6+3+0-4-7-5)^2 \right] \right\} = 7.083$$

Método BQV:

Varianza en bloque tamaño 1:

$$VAR(x) = \left(\frac{2}{7}\right) \left\{ \left[ \frac{1}{2}(8-6)^2 \right] + \left[ \frac{1}{2}(6-3)^2 \right] + \left[ \frac{1}{2}(3-0)^2 \right] + \left[ \frac{1}{2}(0-4)^2 \right] + \left[ \frac{1}{2}(4-7)^2 \right] + \left[ \frac{1}{2}(7-5)^2 \right] \right\} = 7.29$$

Varianza en bloque tamaño 2:

$$VAR(x) = \left(\frac{4}{7}\right) \left\{ \left[ \frac{1}{4}(8+6-3-0)^2 \right] + \left[ \frac{1}{4}(6+3-0-4)^2 \right] + \left[ \frac{1}{4}(3+0-4-7)^2 \right] + \left[ \frac{1}{4}(0+4-7-5)^2 \right] \right\} = 39.143$$

Varianza en bloque tamaño 3:

$$VAR(x) = \left(\frac{6}{7}\right) \left\{ \left[ \frac{1}{6}(8+6+3-0-4-7)^2 \right] + \left[ \frac{1}{6}(6+3+0-4-7-5)^2 \right] \right\} = 12.143$$

Método PQV:

Varianza en bloque tamaño 1:

$$VAR(x) = \left[ \frac{1}{(7-1)} \right] \left\{ \left[ \frac{1}{2}(8-6)^2 \right] + \left[ \frac{1}{2}(6-3)^2 \right] + \left[ \frac{1}{2}(3-0)^2 \right] + \left[ \frac{1}{2}(0-4)^2 \right] + \left[ \frac{1}{2}(4-7)^2 \right] + \left[ \frac{1}{2}(7-5)^2 \right] \right\} = 4.25$$

Varianza en bloque tamaño 2:

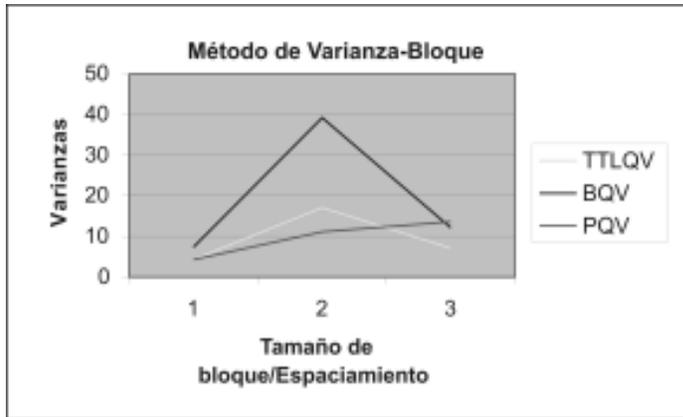
$$VAR(x) = \left[ \frac{1}{(7-2)} \right] \left\{ \left[ \frac{1}{2}(8-3)^2 \right] + \left[ \frac{1}{2}(6-0)^2 \right] + \left[ \frac{1}{2}(3-4)^2 \right] + \left[ \frac{1}{2}(0-7)^2 \right] + \left[ \frac{1}{2}(4-5)^2 \right] \right\} = 11.2$$

Varianza en bloque tamaño 3:

$$VAR(x)B = \left[ \frac{1}{(7-3)} \right] \left\{ \left[ \frac{1}{2}(8-0)^2 \right] + \left[ \frac{1}{2}(6-4)^2 \right] + \left[ \frac{1}{2}(3-7)^2 \right] + \left[ \frac{1}{2}(0-5)^2 \right] \right\} = 13.625$$

Nota: Aunque para este se puede trabajar con tamaño 4, es recomendable que se haga de igual tamaño que los anteriores.

Interpretación: Se presenta agregación en el comportamiento de la orquídea evaluada, con base en la varianza-pico el tamaño de bloque espaciamento tipo 2. Es decir, las orquídeas presentan un comportamiento de tipo gregario, principalmente cuando se realiza el análisis en pares de cuadrantes contiguos (ver gráfica)



### 3.3 Métodos de distancia

Sirven para determinar el patrón espacial de las poblaciones, con base en la distancia de los datos obtenidos. Los índices son derivados de las distancias entre puntos e individuos y se realizan pruebas estadísticas para la hipótesis del patrón espacial sobresaliente.

Este tipo de análisis es más eficiente que el muestreo cuadrático cuando los individuos están esparcidos y ampliamente dispersos.

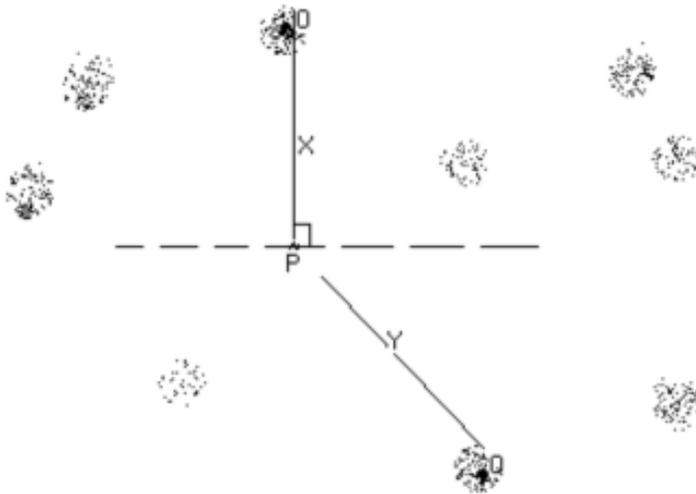
Este muestreo requiere dos distancias que son medidas en cada punto de muestreo localizado en una comunidad:

- La distancia desde el punto de muestreo al individuo más cercano.
- La distancia desde ese individuo al vecino más cercano.

Aquí se trabaja con otro índice de dispersión basado solamente en distancias punto a individuo y se procede de la siguiente manera:

- Los puntos muestrales son seleccionados en una rejilla regular o aleatoria dentro de la comunidad y dos distancias son medidas en cada punto aleatorio.

La distancia  $X$  es medida en cualquier dirección.



- Una línea perpendicular a la línea  $OP$  y se calcula la distancia  $Y$ .

Si el patrón de individuos de una población a través de un área continua es completamente aleatorio, entonces con un muestreo de distancia el cuadrado esperado de distancias punto a individuo será aproximadamente igual a la mitad del cuadrado esperado de T-cuadrado distancias al vecino más cercano:

$$E(x^2) = \frac{1}{2} E(y^2)$$

Sin embargo, si el patrón de individuos es agregado, entonces el cuadrado esperado de las distancias punto a individuo será más grande que la mitad del cuadrado esperado de las distancias al vecino más cercano:

$$E(x^2) > \frac{1}{2} E(y^2)$$

Entonces el patrón uniforme es:

$$E(x^2) = \frac{1}{2} E(y^2)$$

Es de resaltar que las distancias son cuadráticas porque las medidas de distancia son a través de un espacio bidimensional.

-Índice de patrón espacial T-cuadrado:

De las distancias de muestreo T-cuadrado obtenidas, se puede obtener un índice de patrón espacial (C), como una relación de distancias cuadráticas punto a individuo X y las distancias cuadráticas de individuo al vecino más cercano Y, así:

$$C = \frac{\sum_{i=1}^N \left[ \frac{X_i^2}{\left( X_i^2 + \frac{1}{2} Y_i^2 \right)} \right]}{N}$$

Donde N es el número total de puntos muestrales.

El valor de C es aproximadamente la mitad para patrones aleatorios, significativamente menores que la mitad para patrones uniformes y significativamente más grandes para patrones agregados.

$C \cong \frac{1}{2}$  Patrón aleatorio

$C < \frac{1}{2}$  Patrón uniforme

$C > \frac{1}{2}$  Patrón agregado

Para evaluar la significancia de cualquier variación de C desde la mitad, se calcula con un valor Z como:

$$Z = \frac{C - 0,5}{\sqrt{\frac{1}{12N}}}$$

La significancia estadística de Z desde que C sea obtenida por una tabla de probabilidad de Z bajo una significancia del 5% ( $Z=1,96$ ).

-Índice de dispersión de distancia (I):

Es considerada una prueba poderosa para el análisis de patrón espacial, dada una muestra de N puntos con distancias  $X_i$ , desde el i-ésimo punto al individuo más cercano, entonces el índice de dispersión de distancia (I) está definido como la relación de la suma de cuadrados de las distancias cuadradas:

$$I = (N + 1) * \frac{\sum_{i=1}^N (x_i^2)^2}{\left[ \sum_{i=1}^N (x_i)^2 \right]^2}$$

$I=2$  Patrón aleatorio

$I<2$  Patrón uniforme

$I>2$  Patrón agregado

Se ha demostrado que I tiende a la normalidad cuando  $N=100$ , el cual provee la base

para el valor de Z así:

$$Z = \frac{I - 2}{\sqrt{\frac{4 * (N - 1)}{(N + 2) * (N + 3)}}}$$

Este valor de Z es comparado con una tabla de valores críticos de la distribución normal Standard, para obtener el nivel de significancia de cualquier desviación desde la aleatoriedad.

**NOTA:** El muestreo de distancia provee un método eficiente de análisis rápido del patrón espacial de individuos que son reportados a través de una comunidad continua, especialmente cuando los individuos se distribuyen fácilmente (árboles en un bosque, por ejemplo).

Las distancias pueden ser medidas a partir de puntos muestrales seleccionados de una rejilla regular, sin que los índices analizados con antelación pierdan poder.

Ejemplo:

Al analizar la distribución de plantas de una especie x a través de transectos, se determinó la distancia entre puntos de observación y la planta de interés, además de la distancia entre la planta de interés a su vecina más cercana obteniéndose lo siguiente en 10 puntos tomados:

x	y
5.1	6.3
5.3	6.4
6.1	6.2
1.5	5.5
1.8	5.3
1.9	5.2
2.7	5.0
3.9	5.8
4.1	4.0
4.5	4.2

Donde:

x: Distancia del punto a la planta más cercana.

y: Distancia de la planta a su vecina más cercana.

Se debe establecer análisis estadísticos de distancia entre individuos e índices de patrón espacial.

Se realiza así:

Se calcula la distancia promedio del punto a la planta más cercana,

$$\bar{x} = \frac{5.1 + 5.3 + \dots + 4.5}{10} = 3.69$$

Se calcula la distancia promedio de la planta a su vecina más cercana,

$$\bar{y} = \frac{6.3 + 6.4 + \dots + 4.2}{10} = 5.39$$

Seguidamente se calcula el índice T-cuadrado de patrón espacial ( C ),

$$C = \frac{\left[ \frac{5.1^2}{\left( \frac{5.1^2 + 6.3^2}{2} \right)} \right] + \left[ \frac{5.3^2}{\left( \frac{5.3^2 + 6.4^2}{2} \right)} \right] + \dots + \left[ \frac{4.5^2}{\left( \frac{4.5^2 + 4.2^2}{2} \right)} \right]}{10} = 0.46$$

Dado que  $C=0.46$ , presente una tendencia de patrón aleatorio.

Se calcula la significancia de C, con base en la distribución normal standard ( Z ),

$$Z = \frac{0.45 - 0.5}{\sqrt{\frac{1}{12 * 10}}} = -0.43$$

Como  $Z < Z_{\alpha/2}$  (1,96) Se acepta la hipótesis que, por las distancias entre puntos e individuos, se presenta un patrón aleatorio entre plantas.

Se calcula el índice de dispersión de distancia ( I ),

$$I = (10 + 1) * \frac{(5.1^2)^2 + (5.3^2)^2 + \dots + (4.5^2)^2}{(5.1^2 + 5.3^2 + \dots + 4.5^2)} = 1.66$$

Indica una tendencia no marcada a la aleatoriedad.

Se calcula la significancia de I, con base en Z, así:

$$Z = \frac{1.66 - 2}{\sqrt{\frac{4 * (10 - 1)}{(10 + 2)(10 + 3)}}} = -0.71$$

Como  $Z < Z_{\alpha/2}$  (1,96) Se ratifica la hipótesis que, por las distancias entre puntos e individuos, se presenta un patrón aleatorio entre plantas.

## **4. RELACIONES DE ABUNDANCIA DE ESPECIES.**

### **Antecedentes.**

La abundancia de especies gira en torno a la posible variabilidad, generando diferentes tipos de preguntas como: ¿cuáles son las abundancias relativas de las especies?, ¿cuántas especies son raras?, cuáles son comunes?

Las relaciones de abundancia permiten el análisis de estabilidad de comunidades bióticas, proceso evolutivo, entre otros.

La abundancia de especies se basa usualmente en el número de individuos por especie, en forma de distribución de frecuencia, en forma de número total de especies en una comunidad, riqueza y abundancia relativa de las mismas.

Whittaker (1972), citado por Ludwig y Reynolds (1988), definió 3 niveles distintos de diversidad para ecologistas:

1. Diversidad dentro de los hábitats
2. Diversidad entre hábitats
3. Diversidad a gran escala, que comprende las 2 anteriores.

Un amplio rango de juego de datos es posible, sin embargo la sábana (ver página 6, ambas tablas) es una herramienta muy adecuada para la recolección de la información.

### **4.1 Modelos de Distribución**

Cuando se obtiene una gran muestra de abundancia de especies vegetales o animales a partir de una comunidad, se pueden manejar estos datos de diversas maneras para ayudar a examinar las relaciones entre abundancia y el número de especies que la tienen, ya sea en forma de distribución de frecuencia o a través de del ordenamiento de la comunidad desde las especies más abundantes a las menos abundantes.

Estas evaluaciones son útiles al estudiar nichos ecológicos; además, se han planteado tres distribuciones teóricas que son:

1. Distribución geométrica: Es el resultante de una comunidad relativamente pobre de especies, donde un factor ambiental es extremadamente importante para la supervivencia de las mismas (p.ej: temperatura), de donde surge la hipótesis que es posible que una sola especie es dominante y posee una gran fracción de los recursos para vivir, la segunda especie más exitosa posee una más pequeña fracción, y así sucesivamente. Esta hipótesis se conoce como hipótesis de posesión de nicho.
2. Distribución barra rota (broke stick): Cuando el factor ambiental más importante no afecta a las especies y todas utilizan la fracción de recursos sin traslaparse unas con otras.
3. Distribución Lognormal: Cuando existe una gran colección de especies, su abundancia relativa es el producto de la interacción de muchos factores más o menos independientes, generándose la distribución mencionada, la cual está asociada al teorema del límite central.

La importancia de entender los diferentes tipos de distribución espacial radica en que permite la detección de patrones o tendencias que pueden ser útiles para generar hipótesis evaluables acerca de la organización de las comunidades.

Preston (1948, 1962), citado por Ludwig y Reynolds (1988), introdujo el uso de la distribución lognormal en la ecología, usando logaritmos en base 2 de tales clases u octavas y representa un doblamiento de la clase de abundancia previa.

Dicha distribución Lognormal es dada por:

$$S(R) = S_0 * e^{(a^2 R^2)}$$

donde:

$S(R)$  : Número de especies en el R-ésimo octavo

$S_0$  : Estimativo del número de especies en el octavo modal (el octavo con mayor número de especies).

$a$  : Parámetro, medida inversa de la amplitud de la distribución, p. ej:

$$a = \frac{1}{2\sigma}, \text{ donde } \sigma \text{ es la desviación standard.}$$

Paso 1. Distribución de frecuencia observada: Los datos observados son arreglados en la forma de una distribución de frecuencia registrando el número de especies en cada clase u octava y siguiendo la convención de Preston, con base en la distribución lognormal, R para el i-ésimo octavo es dado por:

$$R = \text{Log}_2 \left( \frac{N_i}{N_0} \right)$$

donde:

$N_i$  : Abundancia de especies en el i-ésimo octavo.

$N_0$  : Abundancia de especies en el octavo modal.

Mientras que cada octavo representa un doblamiento de abundancia, la relación  $N_i/N_0$  para los octavos sucesivos a la derecha del modal será 2, 4, 8, 16, 32, etc y R será 1, 2, 3, 4, 5, respectivamente; para el octavo a la izquierda del modal, la relación respectiva será 0.5, 0.25, 0.125, etc y R será -1, -2, -3, etc.

Un ejemplo de lo anterior se muestra así:

Octavo	$N_i / N_0$	$R$	$R^2$	$S(R)$	$\text{Ln } S(R)$
1	0.125	-3	9	$S(R = -3)$	$\text{Ln } S(R = -3)$
2	0.25	-2	4	$S(R = -2)$	$\text{Ln } S(R = -2)$
3	0.5	-1	1	$S(R = -1)$	$\text{Ln } S(R = -1)$
4	1	0	0	$S(R = 0)$	$\text{Ln } S(R = 0)$
5	2	1	1	$S(R = 1)$	$\text{Ln } S(R = 1)$
6	4	2	4	$S(R = 2)$	$\text{Ln } S(R = 2)$
7	8	3	9	$S(R = 3)$	$\text{Ln } S(R = 3)$
8	16	4	16	$S(R = 4)$	$\text{Ln } S(R = 4)$

La columna  $R^2$  es dada siempre y cuando aparezca la ecuación ;  
 $S(R) = S_0 * e^{(-a^2 R^2)}$  además, los logaritmos naturales de  $S(R)$  son usados para estimar los parámetros del modelo lognormal en el paso 2.

La curva lognormal tiende a infinito en el eje de las octavas (eje x) en ambas direcciones. El área bajo la curva es de manera teórica el universo que es muestreado en campo, que se entiende como el número total de especies bajo observación.

Para muestras pequeñas, solamente es visible el lado derecho de la curva.

A medida que aumenta el tamaño de la muestra y posiblemente puedan incluirse las especies más raras de la comunidad evaluada , el lado izquierdo de la curva tiende a proyectarse y ser visible.

Paso 2. Estimación de parámetros: La distribución lognormal se fundamenta en 2 parámetros,  $S_0$  y  $a$

Aproximación para el parámetro  $a$  :

$$a = \sqrt{\frac{\text{Ln}\left[\frac{S_0}{S(R_{max})}\right]}{R^2_{max}}}$$

donde:

$S_0$  : Número observado de especies en el octavo modal.

$S(R_{max})$ : Número observado de especies en el octavo más distante del modal.

La ecuación trabaja cuando los datos incluyen observaciones de al menos 3 o 4 octavos distantes de la moda. Es de agregar que, si el octavo más distante es el mismo en ambas direcciones (p. ej: -6, 6), entonces se calcula un valor para ambos  $S(R_{max})$  y se promedia posteriormente.

El parámetro de estimación  $S_0$  se puede estimar a partir de:

$$S_0 = e^{(\overline{\text{Ln}S(R)} + a^2 \overline{R^2})}$$

Donde:

$\overline{\text{Ln}S(R)}$  : Es el promedio de los logaritmos del número observado de especies por octavo.

$a$  : Parámetro estimativo (Ecuación mostrada con antelación).

$\overline{R^2}$  : Es el promedio de los  $R^2$ .

El programa LOGNORMAL permite una sustitución iterativa de diferentes valores para  $a$  y  $S_0$  dentro de la ecuación principal, hasta las desviaciones entre las observaciones y el número esperado de especies en los octavos (paso 3) son minimizados.

Paso 3. Cálculo de las frecuencias esperadas: Usando las estimaciones para  $a$  y  $S_0$ , las frecuencias esperadas lognormal son calculados usando la ecuación principal y la aptitud del modelo se analiza con

base en la distribución  $\chi^2$ , donde los grados de libertad son iguales al número de clases octavas menos 2.

Es de resaltar que el modelo lognormal puede indicar que un sistema está en equilibrio.

Ejemplo:

Se está evaluando la abundancia de insectos en cierto bosque y se observó en 10 cateos que:

Nº de Individuos/ spp(*)	1	2	3	4	5	8	16	24	45	80	188
Frecuencia observada de la spp(**)	26	20	15	8	2	4	6	1	5	4	91

(\*) Se organiza de menor a mayor

(\*\*) Número de especies (spp)

Existen en total en los 10 cateos, 188 individuos provenientes de 91 especies.

Para el cálculo de la distribución Lognormal, se realiza una distribución de frecuencia así:

0-1 (+1), 1-2 (+1), 2-4 (+2), 4-8 (+4), 8-16 (+8), 16-32 (+16), 32-64 (+32), 64-128 (+64).

Que son las respectivas octavas.

Se detiene aquí dado que el máximo número de individuos encontrado es 80 y está dentro del último intervalo (64-128, respectivamente).

Posteriormente, se calculan los valores observados por octava:  $S(R)$ ;

$$Octava(0-1) = \frac{26}{2} = 13$$

$$Octava2(1 - 2) = 13_{+deoctava1} + \frac{20}{2} = 23$$

$$Octava3(2 - 4) = 10_{+deoctava2} + 15 + \frac{8}{2} = 29$$

$$Octava4(4 - 8) = 4_{+deoctava3} + 2 + \frac{4}{2} = 8$$

$$Octava5(8 - 16) = 2_{+deoctava4} + \frac{6}{2} = 5$$

$$Octava6(16 - 32) = 3_{+deoctava5} + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 1_{+posición24} + 0 + \dots + 0_{+posición32} = 4$$

$$Octava7(32 - 64) = 0_{+deoctava6} + 0 + \dots + 5_{+posición45} + \dots + 0_{+posición64} = 5$$

$$Octava8(64 - 128) = 0_{+deoctava7} + 0 + \dots + 4_{+posición80} + 0_{+posición128} = 4$$

Obteniéndose la siguiente distribución:

<b>Octava</b>	Nºindividuos /spp	R	R <sup>2</sup>	S(R)Observado	LnS(R)
1	0-1	-2	4	13	2.56
2	1-2	-1	1	23	3.14
3	2-4	0	0	29 (*)	3.37
4	4-8	+1	1	8	2.08
5	8-16	+2	4	5	1.61
6	16-32	+3	9	4	1.39
7	32-64	+4	16	5	1.61
8	64-128	+5	25	4	1.39

(\*) como 29 es la frecuencia observada más alta se denomina frecuencia modal, siendo R igual a 0 por ser el más alto, hacia arriba en la distribución son negativos, -1, -2 y hacia abajo son positivos, +1, +2, +3, +4 y +5, respectivamente.

Después se estima el parámetro  $a$ , a partir del  $R$  máximo ( $R_{\max}$ ) =+5 y  $S_0$  =29,  $S(R_{\max})$  =4 y  $R^2$  máximo ( $R_{\max}^2$ )=25:

$$a = \sqrt{\frac{\ln\left(\frac{29}{4}\right)}{25}} = 0.28$$

Luego se calculan:

$$\overline{\ln S(R)} = \frac{2.56 + 3.14 + \dots + 1.39}{8} = 2.14$$

$$\overline{R^2} = \frac{4 + 1 + 0 + \dots + 25}{8} = 7.50$$

$$S_0 = e^{2.14 + 0.28^2 * 7.50} = 15.30$$

Después son calculadas las frecuencias esperadas usando  $a=0.28$  y  $S_0$  =29 y 15.30.

Con  $a=0.28$  y  $S_0=29$

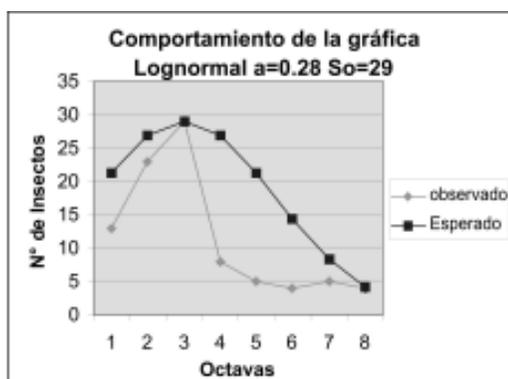
$$S(-2) = 29 * e^{-(0.28)^2 (-2)^2} = 21.19$$

$$S(-1) = 29 * e^{-(0.28)^2 (-1)^2} = 26.81 \text{ y así sucesivamente hasta}$$

$$S(+5) = 29 * e^{-(0.28)^2 (+5)^2} = 4.08$$

Cuya frecuencia es:

Octava	R	S(R) observado	S(R) esperado	$\chi^2$
1	-2	13	21.19	3.17
2	-1	23	26.81	0.54
3	0	29	29	0
4	+1	8	26.81	13.20
5	+2	5	21.19	12.37
6	+3	4	14.32	7.44
7	+4	5	8.27	1.29
8	+5	4	4.08	0
Sumatoria		91	151.67	38.01



Seguido se calcula:

$$S^* = 1.77 \left( \frac{S_0}{a} \right) \rightarrow S^* = 1.77 \left( \frac{29}{0.28} \right) = 183.32 \approx 183 \rightarrow S^* - \sum S(R) = 183 - 91 = 92$$

Indicando que hay 92 especies no observadas en la comunidad evaluada en el bosque bajo estudio.

No obstante, se debe calcular con a=0.28  $S_0=15.30$  de la misma manera que se hizo anteriormente. Es de resaltar que, cuanto más baja sea la sumatoria de  $\chi^2$ , el modelo es más ajustado a la realidad.

Con  $a=0.28$  y  $S_0 = 15.30$

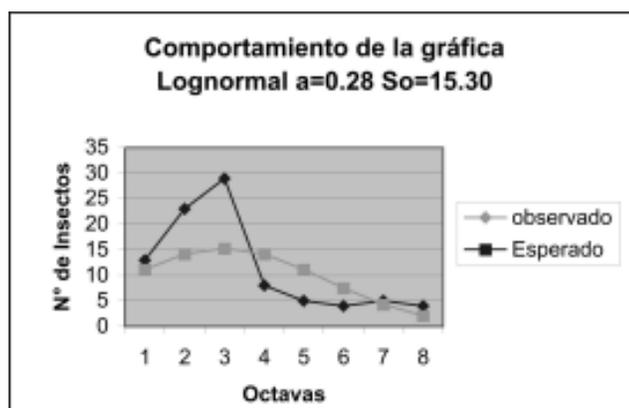
$$S(-2) = 15.30 * e^{-(0.28)^2 (-2)^2} = 11.18$$

$S(-1) = 15.30 * e^{-(0.28)^2 (-1)^2} = 14.15$  y así de manera sucesiva hasta

$$S(+5) = 15.30 * e^{-(0.28)^2 (+5)^2} = 2.16$$

Cuya frecuencia es:

Octava	R	S(R) observado	S(R) esperado	$\chi^2$
1	-2	13	11.18	0.30
2	-1	23	14.15	5.54
3	0	29	15.3	12.27
4	+1	8	14.15	2.67
5	+2	5	11.18	3.42
6	+3	4	7.56	1.68
7	+4	5	4.36	0.09
8	+5	4	2.16	1.57
Sumatoria		91	80.04	27.53



Seguido se calcula:

$$S^* = 1.77 \left( \frac{S_0}{a} \right) \rightarrow S^* = 1.77 \left( \frac{5.30}{0.28} \right) = 96.72 \approx 97 \rightarrow S^* - \sum S(R) = 97 - 91 = 6$$

Indicando que hay 6 especies no observadas en la comunidad evaluada en el bosque bajo estudio.

Además, dado que  $X^2$  para la segunda serie de cálculos es más bajo, se ajusta más a la realidad, debiéndose realizar los análisis respectivos con base en tal serie.

## 4.2 Índices de Diversidad

Estos son utilizados para caracterizan la relación de abundancia de especies en comunidades bióticas.

La diversidad está compuesta por dos componentes distintos:

- Número total de especies
- Igualdad (Como los datos de abundancia están distribuidos entre las especies).

Para entender los conceptos anteriores se plantea el siguiente ejemplo: Si en un bosque existen 20 especies y el 80% de ellos pertenecen a una sola especie, indica que la igualdad es baja. En caso contrario, si cada especie tuviera el 5% de los individuos, la igualdad se considera máxima.

Los índices están en función de la riqueza de especies y la igualdad y los índices que combinan ambos componentes son llamados índices de diversidad. Aunque, se ha criticado dicha combinación de ambos, dado que tiende a confundir el número de variables que caracterizan la estructura de una comunidad: el número de especies, la abundancia relativa de las mismas y la homogeneidad y tamaño del área muestreada.

Índices de riqueza: El índice de riqueza de especies (S) da la apariencia de ser ambiguo pero depende del tamaño de la muestra.

Sin embargo, los índices más conocidos son:

Índice de Margalef (1958):

$$R_1 = \frac{S-1}{\ln(n)}$$

Índice de Menhinick (1964):

$$R_2 = \frac{S}{\sqrt{n}}$$

Ambos deben ser usados como índices de riqueza exclusivamente; además, una alternativa para determinar dichos índices, es usar conteos directos del número de especies en muestras de igual tamaño.

En situaciones donde los tamaños de muestra no son iguales (es probable que sea lo más común), existe un método estadístico conocido como "rarefraction" (Hulbert (1971), citado por Ludwing y Reynolds (1988)), el cual asume a las diferencias existentes en el tamaño de la muestra, donde el número de especies que pueden ser esperadas en una muestra de n individuos(denotados por E(Sn))a partir de una población total de N individuos distribuidos entre S especies.

Se denota de la siguiente manera:

$$E(Sn) = \sum_{i=1}^s \left\{ 1 - \left[ \frac{\binom{N-ni}{n}}{\binom{N}{n}} \right] \right\}$$

donde:

$ni$  : número de individuos de la i-ésima especie

En otras palabras, la anterior ecuación calcula el número esperado de especies en una muestra aleatoria de tamaño n, como la suma de las probabilidades que cada especie será incluida en la muestra.

Este método es de gran utilidad cuando son llevados a cabo censos de especies en comunidades.

Ludwing y Reynolds (1988) afirman que los métodos "rarefraction" son preferidos a los índices de riqueza simples cuando los tamaños de muestra de la comunidad difieren.

Índice de diversidad: Los índices de diversidad incorporan tanto la igualdad (uniformidad), como la riqueza (abundancia) de especies dentro de un solo valor. Es por esto que, Peet (1974), citado por Ludwing y Reynolds (1988), define dichos índices como índices de heterogeneidad.

Los números de diversidad de Hill (1973), citado por Ludwing y Reynolds (1988) son:

N0: Número total de especies, independientemente de la abundancia.

N2: Número de especies muy abundantes.

N1: Número de especies abundantes (intermedio entre N0 y N2).

No obstante, dos índices son necesitados para calcular los anteriores números de diversidad: El índice de Simpson ( $\lambda$ ) / el índice de Shannon ( $H'$ ) los cuales se describen de la siguiente manera:

Índice de Simpson:

$$\lambda = \sum_{i=1}^S P_i^2$$

donde  $P_i$  es la abundancia proporcional de la  $i$ -ésima especie, dada por:

$$P_i = \frac{n_i}{N} \quad i=1, 2, 3, \dots, S$$

$n_i$ : Número de individuos de la  $i$ -ésima especie

$N$ : Número total conocido de individuos para todas las  $S$  especies en la población.

Dichos índice varía entre 0 y 1 ( $0 \leq \lambda \leq 1$ ), y está dado por la probabilidad que dos individuos aleatorios de una población pertenezcan a la misma especie. Cuando la probabilidad es alta para ambos individuos si pertenecen a la misma especie, entonces la diversidad de la muestra tiende a bajar.

La ecuación planteada anteriormente solo se aplica a poblaciones finitas y los individuos han sido contados ( $n = N$ , donde  $n$  es el número total de individuos en la muestra y  $N$  es el número total de individuos en la población).

Aunque, Simpson (1949) desarrolló un estimador para muestreo a partir de una población infinita ( $\lambda$ ):

$$\hat{\lambda} = \sum_{i=1}^s \frac{ni(ni-1)}{n(n-1)}$$

Índice de Shannon ( $H'$ ): Este índice se basa en el grado promedio de incertidumbre a que especie pertenece un individuo tomado en forma aleatoria de una colección de  $S$  especies y  $N$  individuos.

Este promedio de incertidumbre se incrementa cuando se incrementa el número de especies.

El índice de Shannon posee dos propiedades:

- $H' = 0$  cuando solamente hay una especie en la muestra.
- $H'$  es el máximo solamente cuando todas las  $S$  especies están representadas por el mismo número de individuos, lo cual es una distribución perfecta de abundancia.

La ecuación es:

$$H' = -\sum_{i=1}^s (pi * Ln(pi))$$

$H'$  : Promedio de incertidumbre por especie en una comunidad infinita hecha por encima de  $S'$  especies con abundancias proporcionales conocidas  $p_1, p_2, p_3, \dots, p_{S'}$ .

$S'$  y  $p_i$  son parámetros poblacionales y en la práctica  $H'$  es estimado a partir de una muestra así:

$$\hat{H}' = -\sum_{i=1}^S \left[ \left( \frac{n_i}{n} \right) \ln \left( \frac{n_i}{n} \right) \right]$$

Donde  $n_i$  es el número de individuos pertenecientes a la  $i$ -ésima de  $S$  especies en la muestra y  $n$  es el número total de individuos en la muestra.

### 4.3 Índices de Uniformidad

Cuando todas las especies en una muestra son igualmente abundantes puede intuirse que la uniformidad es máxima y decrece cuando la abundancia relativa de las especies divergen de dicha uniformidad.

Hulbert (1971), citado por Ludwig y Reynolds (1988), notó dicha uniformidad así:

$$V' = \frac{D}{D_{max}} \quad \text{o como} \quad V = \frac{D - D_{min}}{D_{max} - D_{min}}$$

$D$ : Es un índice de diversidad observado

$D_{min}$ - $D_{max}$ : Valores mínimo y máximo

Aunque se pueden generar diversos índices, los más utilizados son cinco, que son:

Índice 1 ( $E_1$ ): Es probablemente el índice de uniformidad más comúnmente usado por los ecólogos.

$$E_1 = \frac{H'}{\ln(S)} = \frac{\ln(N_1)}{\ln(N_0)}$$

este es el  $J'$  de Pielou (1975, 1977), citado por Ludwig y Reynolds (1988), el cual se expresa como  $H'$  relativa al máximo valor que  $H'$  puede obtener cuando todas las especies en la muestra son perfectamente uniformes con base en un individuo por especie.

Índice 2 (E2): Sheldon(1969), citado por Ludwig y Reynolds (1988), propuso este índice como la forma exponencial de E1, así:

$$E2 = \frac{e^{H'}}{S} = \frac{N1}{N0}$$

Índice 3 (E3): Con base en E2, es el índice de uniformidad propuesto por Heip (1974), citado por Ludwig y Reynolds (1988), el cual se denota de la siguiente manera:

$$E3 = \frac{e^{H'} - 1}{S - 1} = \frac{N1 - 1}{N0 - 1}$$

Índice 4 (E4): Es la relación entre N2 a N1, como índice de diversidad:

$$E4 = \frac{1/\lambda}{e^{H'}} = \frac{N2}{N1}$$

Esta es la relación que se aplica para el número de especies muy abundantes a especies abundantes. Parte de la discusión planteada acerca de cómo una comunidad decrece, que es como una especie tiende a dominar, ambos N2 y N1 tienden a 1, por lo tanto E4 converge hacia dicho valor.

Índice 5 (E5): Es conocido como la relación modificada de Hill, el cual se aproxima a cero cuando una sola especie llega a ser cada vez más dominante en una comunidad.

$$E5 = \frac{(1/\lambda)^{-1}}{e^{H'} - 1} = \frac{N2 - 1}{N1 - 1}$$

Es de resaltar que, cuando en una muestra aparece una especie "rara", donde tal muestra contiene pocos individuos (bajo número de especies) cambia de manera marcada a E1, lo cual sucede de manera similar para E2 y E3. En contraste a lo anterior, E4 y E5 no se ven afectados por la riqueza de especies.

El número actual de especies presentes en la comunidad ( $S^*$ ) se conoce cuando se calculan los índices E1, E2 o E3.

Sin embargo, en la práctica,  $S^*$  es estimado a través del número de especies presentes en la muestra, aunque en ningún momento debe ser subestimado.

En cuanto a los índices E4 y E5, estos permanecen relativamente constantes con respecto a las variaciones del muestreo dado que son independientes del tamaño de la muestra, porque dichos índices son calculados como relaciones, donde el tamaño de la muestra ( $S$ ), está tanto en el numerador como en el denominador, cancelando el impacto del número de especies en la muestra.

Ejemplo:

Fueron analizadas 5 especies de coleópteros en un bosque, donde se realizó el conteo de ellos, encontrándose:

Género 1: 15 individuos

Género 2: 20 individuos

Género 3: 30 individuos

Género 4: 15 individuos

Género 5: 10 individuos

Se calculan los diferentes índices así:

Índice de Margalef:

$$R_1 = \frac{(5-1)}{\ln 90} = 0.8889266$$

Índice de Menhinick:

$$R_2 = \frac{5}{\sqrt{90}} = 0.5270463$$

Estos indican que no hay igual distribución entre el número de especies y sus abundancias relativas.

Se calculan los índices de diversidad:

Índice de Simpson ( $\lambda$ ):

$$\lambda = \frac{[(15 * (15 - 1)) + (20 * (20 - 1)) + \dots + (10 * (10 - 1))]}{90 * (90 - 1)} = 0.2072409$$

Dado que el índice de Simpson es igual a 0.2072409, indica que al tomar en forma aleatoria un par de individuos, la probabilidad que ambos sean de la misma especie es del 20.72409%, tendiendo a una diversidad alta.

Índice de Shannon ( $H'$ ):

$$H' = - \left[ \left( \left( \frac{15}{90} \right) * \ln \left( \frac{15}{90} \right) \right) + \left( \left( \frac{20}{90} \right) * \ln \left( \frac{20}{90} \right) \right) + \dots + \left( \left( \frac{10}{90} \right) * \ln \left( \frac{10}{90} \right) \right) \right] = 1.567057$$

Muestra una tendencia hacia la diversidad de especies (dado que se aleja de cero).

Luego se calculan  $N_1$  y  $N_2$  así:

$N_1$  con base en el índice de Shannon:

$N_1 = e^{H'} = e^{1.567057} = 4.792521 \approx 5$ , lo cual explica que existen 5 especies abundantes.

$N_2$  con base en el índice de Simpson:

$$N_2 = \frac{1}{\lambda} = \frac{1}{0.2072409} = 4.825302$$

Dado que  $N_2 > 1$  indica tendencia a la diversidad de especies.

Posteriormente se calculan otros índices de uniformidad:

$$E_1 = \frac{\text{Ln}(4.792521)}{\text{Ln}(5)} = 0.973667 \text{ (Es el más recomendable de } E_1, E_2, E_3 \text{ )}$$

$$E_2 = \frac{4.792521}{5} = 0.9585041$$

$$E_3 = \frac{4.792521 - 1}{5 - 1} = 0.9481301$$

Los 3 índices indican el máximo valor que  $H'$  puede obtener cuando todas las especies en la muestra están distribuidas de manera perfectamente uniforme, con base en un individuo por especie.

$$E_4 = \frac{4.825302}{4.792521} = 1.00684$$

Indica que existe una relación equilibrada entre el número de especies muy abundantes y aquellas abundantes, tendiendo hacia la uniformidad, dado que es un valor cercano a uno.

$$E_5 = \frac{4.825302 - 1}{4.792521 - 1} = 1.00864$$

Como este índice es mayor a cero, indica que no existe una especie dominante en la comunidad evaluada (cuando  $E_5 = 0$ , denota que existe una especie que es dominante).

## **5 AFINIDAD DE ESPECIES**

### **Antecedentes.**

Las comunidades ecológicas están compuestas por un número de especies que coexisten, por lo cual se busca su examen y análisis, dado que las especies podrían estar utilizando recursos comunes.

Las medidas de afinidad son el grado de cómo las especies se traslapan en el uso de recursos comunes.

Este traslape es definido en términos de varias porciones del nicho de una especie que es ocupado por otras.

Los estudios de nichos están basados en tales atributos de especies tales como dieta, preferencia a ciertos hábitats, sincronización de actividades como pastoreo, entre otros.

La afinidad es medida al interior del análisis de la ocurrencia de las especies (al evaluar una serie de unidades de muestreo) que podrían ser más grandes o menos de lo que podría esperarse si estas fueran independientes.

La afinidad se basa en datos de presencia/ausencia; además, si la muestra contiene mediciones cuantitativas de abundancia de especies, es posible determinar la covariación en las abundancias entre especies. Los estudios de afinidad se pueden realizar por unidades de muestreo o por especies.

### **5.1 Índices de traslape de Nicho**

Al estudiar los traslapes de nichos surge un cuestionamiento: ¿Cómo coexisten las especies que utilizan recursos comunes en una comunidad?.

Aquellas especies que comparten un recurso común, poseen entonces un alto grado de traslape y aquellas con bajo compartir de recursos tienen bajo traslape.

El nicho es considerado como la posición de las especies en una comunidad, incluyendo uso de recursos de comida, tiempo de actividad, localización, modo de interacción con otras especies y así sucesivamente. Según Petraitis (1979), citado por Ludwig y Reynolds (1988), el traslape de nichos es determinante en el estudio de diversidad de especies y estructura de la comunidad, aunque puede estar supeditado a la subjetividad.

Sin embargo el autor anteriormente mencionado desarrolló una medición de traslape, basada en el criterio que la probabilidad del uso de los recursos para una especie es idéntico que para otras especies.

Los índices de traslape generales y específicos se calculan de diferentes maneras así:

Con base en el uso relativo de ciertos recursos por cada especie esta determinado por su curva de uso.

Índice de Levins (LO), mide el grado de traslape de dos curvas de uso, de la especie uno con la especie dos, que es dada por:

$$LO_{1,2} = \frac{\sum_j [(P_{1j})(P_{2j})]}{\sum_j (P_{1j}^2)}$$

Tabla 3: En forma de resumen de los términos de cálculo de las mediciones general y específica del traslape se plantea aquí para tres recursos (r=3) por dos especies (s=2). Se suscribe la i-ésima especie y el j-ésimo recurso. Modificado por Petraitis (1985), citado por Ludwig y Reynolds (1988):

		Clases de Recursos			
Especie		1	2	3	Total
Altitud	1	n1,1	n 1,2	n1,3	N1
	2	n2,1	n2,2	n2,3	N2
Altitud Total		t1	t2	T3	T

Proporción	1	P1,1	P1,2	P1,3
	2	P2,1	P2,2	P2,3
Proporciones combinadas		C1	C2	C3

$$N_i = \sum_{j=1}^r (n_{ij}) \quad t_j = \sum_{i=1}^s (n_{ij}) \quad T = \sum_{j=1}^r (t_j) \quad P_{ij} = \frac{n_{ij}}{N_i} \quad C_j = \frac{t_j}{T}$$

La diferencia que se genera es el resultado de cómo la cantidad de traslape entre 2 especies es estandarizado, el denominador del índice de Levins está planteado en términos de amplitud de la i-ésima especie:

$$B_i = \frac{1}{\sum_j (p_{ij}^2)}$$

Las amplitudes de las especies son los usos relativos por otras especies (especies utilizadas para ciertos fines).

### 5.1.1. Traslape específico (SO)

Este traslape está basado en la comparación de las curvas de uso de los recursos y se entiende como la cantidad de traslape específico por la especie i dentro de la especie k y es la probabilidad que la curva de uso de la especie i pudo haber sido marcada a partir de la curva de uso de la especie k.

Usando el tipo de datos de especies-recursos ilustrados en la tabla 3, es posible calcular la probabilidad de la especie  $i$  con respecto al uso de recursos (por ejemplo:  $ni1, ni2, ni3$ ) que pudieron ser trazados a partir de la curva de uso de la especie  $k$  (por ejemplo:  $Pk1, Pk2, Pk3$ ). Para desarrollar el análisis de traslape específico (SO), se definen ciertos pasos:

1. Establecimiento de la hipótesis: Dos tipos generales de hipótesis pueden ser probadas,  $H_0$ : Que dos especies se traslapan completamente, es decir, que sus curvas de uso son iguales.  $H_0$ : Que el traslape específico por la especie  $i$  dentro de una especie  $k$ , es más grande que el traslape de la especie  $i$  dentro de una segunda especie  $m$ .

2. Cálculo del traslape específico: El traslape específico se da para dos especies así:

$$SO_{1,2} = e^{E_{1,2}} \text{ y } SO_{2,1} = e^{E_{2,1}}$$

respectivamente, donde

$$E_{1,2} = \sum_j (P_{1j} \ln P_{2j}) - \sum_j (P_{1j} \ln P_{1j}) \text{ y } E_{2,1} = \sum_j (P_{2j} \ln P_{1j}) - \sum_j (P_{2j} \ln P_{2j})$$

el cálculo de SO requiere de ambas especies para utilizar todos los recursos (clases de ellos), ya que un uso específico de especies de un recurso es igual a cero, entonces  $\ln(P_{ij})$  es indefinido.

3. Cálculo de las pruebas estadísticas: En primera instancia, para probar la hipótesis nula que el traslape específico de la especie  $i$  sobre la especie  $k$  es completo, calculamos lo siguiente:

$$U_{i,k} = -2N_i * \ln(SO_{i,k})$$

Donde  $U_{i,k}$  es distribuida como una  $\chi^2$  con  $r-1$  grados de libertad, entonces para el traslape de la especie 1 sobre la especie 2 es:

$$U_{1,2} = -2N_1 * \ln(SO_{1,2}) \text{ e inverso para } U_{2,1}$$

Posteriormente, para probar la hipótesis que el traslape específico para la especie  $i$  dentro de la especie  $k$  es más grande que el traslape de la especie  $i$  sobre la especie  $m$ , se calcula la relación logarítmica  $W$  como:

$$W = N_j * \text{Ln} \left( \frac{SO_{i,k}}{SO_{i,m}} \right)$$

Si  $W > 2$ , el traslape de la especie  $i$  sobre la especie  $k$  es más grande que el traslape de  $i$  sobre la especie  $m$ .

### 5.1.2. Traslape General (GO)

La cantidad de traslape general entre 2 especies lo define Petraitis como la probabilidad de las curvas de uso de todas las especies que fueron trazadas desde una curva de uso común, aquí no hay mecanismos para hacer comparaciones entre pares de especies.

El traslape general (GO) es calculado como el promedio pesado de curvas de uso de especies así:

Paso 1. Estado de la hipótesis: La hipótesis a probar es que hay traslape completo de las especies.

Paso 2. Cálculo del traslape general (GO) de la siguiente manera:

$$GO = e^E \text{ donde } E = \frac{\sum_i \sum_j [n_{ij} (\text{Ln} C_j - \text{Ln} P_{ij})]}{T}$$

Paso3. Calcular la prueba estadística:

$$V = -2T * \text{Ln} GO$$

$V$  es distribuido como una variación de  $\chi^2$  con  $(s-1)(r-1)$  grados de libertad. Si  $V$  excede el valor crítico para  $\chi^2$ , por decir, para una significancia del 5%, entonces la hipótesis nula de traslape completo es rechazada.

Ejemplo:

Se está investigando la posible preferencia que tienen 3 especies de lepidópteros en estado larval que coexisten, sobre 4 especies de plantas, fueron realizados los conteos respectivos encontrándose lo siguiente:

Especie(larvas)	Plantas				totales
	1	2	3	4	
1	65	10	50	16	141
2	53	115	93	20	281
3	11	26	12	85	134
Totales	129	151	155	121	556

Cuyas proporciones son:

Especie(larvas)	Plantas				Proporción Total
	1	2	3	4	
1	0.461	0.071	0.355	0.113	
2	0.189	0.409	0.331	0.071	
3	0.082	0.194	0.09	0.634	
Proporciones combinadas	0.232	0.272	0.279	0.218	1

Se calculan los traslapes específicos:

$$E_{1,2} = (0.461 \ln 0.189) + \dots + (0.113 \ln 0.071) - ((0.461 \ln 0.461) + \dots + (0.113 \ln 0.113)) = -0.3641$$

$$SO_{1,2} = e^{-0.3641} = 0.694$$

y de manera sucesiva hasta

$$E_{3,2} = (0.082 \ln 0.189) + \dots + (0.634 \ln 0.071) - ((0.082 \ln 0.082) + \dots + (0.634 \ln 0.634)) = -1.0577$$

$$SO_{1,2} = e^{-1.0577} = 0.347$$

Obteniéndose la siguiente tabla de traslapes específicos:

Especie	1	2	3
1	1	0.694	0.362
2	0.611	1	0.477
3	0.36	0.347	1

Después se calcula U así:

$$U_{1,2} = -2(141) \ln 0.694 = 102.953$$

...

$$U_{3,2} = -2(134) \ln 0.347 = 283.381$$

y se obtiene la siguiente tabla:

Especie	1	2	3
1	xxx	102.953	286.744
2	276.906	xxx	415.471
3	273.91	283.381	xxx

Luego se prueba la hipótesis nula que el traslape específico de una especie sobre otra es completa.

Dicha hipótesis se rechaza cuando los valores calculados son mayores o iguales a:

$$X_{\alpha}^2(GL-1) \Rightarrow X_{0.05}^2(3-1) = 5.99$$

Con base en lo anterior, se rechaza la hipótesis de traslape completo entre especies. Sin embargo, esta prueba no define si no hay traslape o algún traslape, dado que estas posibilidades se pueden presentar.

Los resultados finales se presentan en este cuadro resumen:

Traslapes	SO	U
1,2	0.694	102.953
2,1	0.611	276.906
1,3	0.362	286.744
3,1	0.36	273.91
2,3	0.477	415.471
3,2	0.347	283.381

### *Traslapes Específicos:*

Para hallar los traslapes específicos, se utiliza la relación logarítmica  $W$  (con base en el parámetro dos) para cada par de especies:

$$W_{1,3}^{1,2} = 141 \text{Ln} \left( \frac{0.694}{0.362} \right) = 91.77 > 2$$

El traslape de la especie 1 sobre la especie 2 es más grande que aquel de la especie 1 sobre la especie 3.

$$W_{1,2}^{1,3} = 141 \text{Ln} \left( \frac{0.362}{0.694} \right) = -91.77 < 2$$

Ratifica lo dicho anteriormente.

$$W_{2,3}^{2,1} = 281 \text{Ln} \left( \frac{0.611}{0.477} \right) = 69.57 > 2$$

El traslape de la especie 2 sobre la especie 1 es más grande que aquel de la especie 2 sobre la especie 3.

$$W_{3,2}^{3,1} = 134 \text{Ln} \left( \frac{0.360}{0.347} \right) = 4.92 > 2$$

El traslape de la especie 3 sobre la especie 1 es más grande que aquel de la especie 3 sobre la especie 2.

Se define, con base en el uso del recurso "hojas", que existen 3 tipos de traslapes dominantes (1-2, 2-1 y 3-1): es decir, que de manera específica son estos los tipos de afinidades o de ocurrencias de las 3 especies de lepidópteros.

Traslape General:  
Para especie 1.

$$E = \frac{[65(\text{Ln}0.232 - \text{Ln}0.461) + 10(\text{Ln}0.272 - \text{Ln}0.071) + 50(\text{Ln}0.279 - \text{Ln}0.355) + 16(\text{Ln}0.218 - \text{Ln}0.113)]}{556}$$

$$E = -0.2159$$

$$\hat{GO} = e^{-0.2159} = 0.806$$

Seguidamente se calcula:

$$V = -2(556)\text{Ln}(0.806) = 239.83$$

Valor crítico de  $X^2 \Rightarrow X_{\alpha}^2(3\text{spp} - 1) * (4\text{plantas} - 1) \Rightarrow X_{0.05}^2(2) * (3) \Rightarrow X_{0.05}^2(6) = 12.59$

Dado que  $V_{(239.83)} > X_{\text{tabla}(12.59)}^2$ , con un 95% de confiabilidad se rechaza la hipótesis que existe traslape completo, aunque el índice de traslape General ( $\hat{GO} = 0.806$ ) sea grande.

## 5.2 Asociación Interespecífica

Las interacciones de especies son importantes en la ecología, ya que dentro de una comunidad hay diferentes factores bióticos y abióticos que influyen en la distribución, abundancia y subsecuentemente, las interacciones de las especies, dependiendo en si o no dos especies seleccionan o evitan el mismo hábitat, poseen alguna atracción mutua o repulsión o no poseen interacción, resultando un patrón de asociación interespecífica entre especies y presentan índices para medir el grado de asociación. Estas técnicas están basadas en la variable dicotómica presencia/ausencia de las especies en las unidades de muestreo (UM). En general, una asociación interespecífica se da por:

1. Ambas especies seleccionan o evitan el mismo hábitat o factores ambientales propios del mismo.
2. Ambas especies requieren los mismos factores ambientales tanto bióticos como abióticos.
3. Una o ambas especies tienen afinidad por la otra, en forma de atracción o repulsión.

La detección de la asociación de especies tiene importantes implicaciones ecológicas. Algunos procesos ecológicos que pueden convertirse en una asociación positiva o negativa se observan en la tabla 4 y se describen de la siguiente manera:

*Tabla 4. Procesos ecológicos e interacciones interespecíficas que pueden resultar en asociaciones entre especies de índole positivo o negativo(Schluter (1984),citado por Ludwing y Reynolds (1988)):*

<b>Interacción</b>	<b>Positiva</b>	<b>Negativa</b>
<b>Ninguna</b>	Las especies tienen una relación común a un suplemento de recursos ilimitados.	Las especies tienen diferentes Requerimientos de recursos
<b>Mutualismo</b>	Las especies establecen con otras, Probabilidades de supervivencia.	Los recursos son competidos y son Usados exclusivamente por las especies.
<b>Competencia</b>	Las especies fluctúan al unísono en respuesta a recursos limitados.	La interferencia entre especies produce Exclusión ocasional.
<b>Predación</b>	Los depredadores fluctúan en respuesta positiva a las variaciones de los Depredados.	Las altas densidades de depredadores Producen una depresión local de los Depredados.

La detección del tipo de interacción no provee una comprensión causal de la misma; no obstante, dicha detección de un patrón permite de manera ideal la generación de una hipótesis de posibles factores causales sobresalientes, de los cuales existen otros tipos de estudios.

La asociación de especies posee dos componentes distintos, los cuales son de suma importancia:

1. Prueba estadística de hipótesis que dos especies están asociadas o no en algún nivel de probabilidad predeterminado.
2. La medición del grado de precisión de la asociación.  
El procedimiento para estudiar asociación interespecífica está basado en la presencia/ausencia al interior de una colección de unidades de muestreo, lo cual puede ser representado por datos de índole binario: Presencia=1, Ausencia=0.

Tales unidades de muestreo pueden ser naturales o artificiales; aunque son conocidos los problemas en las unidades en mención con respecto al tamaño y la forma, estos pueden ser superados si las unidades son seleccionadas con base en el tamaño, forma y distribución espacial de las especies bajo estudio.

La unidad de muestreo puede ser lo suficientemente grande para incluir potencialmente al menos un individuo de cada especie y no ser tan grande que una de estas especies sea incluida en cada unidad de muestreo.

Para la prueba de asociación están definidos claramente unos pasos:

Paso 1. Suma de datos: Para cada par de especies A y B se obtiene lo siguiente:

- a: Número de unidades de muestreo donde aparecen ambas especies.
- b: Número de unidades de muestreo donde aparece la especie a, mas no la especie b.
- c: Número de unidades de muestreo donde aparece la especie b, mas no la especie a.
- d: Número de unidades de muestreo donde no aparecen las especies a y b.
- N: Número total de unidades de muestreo,  $N=a+b+c+d$ .

La tabla que se construye se organiza de la manera 2x2 y es a partir de ésta que la prueba y las mediciones se basan en estos datos.

La frecuencia esperada de A (Ocurrencia de A) en las unidades de muestreo se representa como  $F(A)$  y se da por:

$$F(A) = \frac{a+b}{N}$$

Y para B, su frecuencia es de la siguiente manera:

$$F(B) = \frac{a+c}{N}$$

Se asume además que, ambas especies han estado presentes en al menos una unidad de muestreo en la colección de datos, que es  $F(A)$  y  $F(B)$  más grandes que cero.

Tabla 5. Tabla de contingencia 2x2 para presencia/ausencia para dos especies:

		Especie B		
		Presencia	Ausencia	
Especie A	Presencia	a	b	m=a+b
	Ausencia	c	d	n=c+d
		r=a+c	s=b+d	N=a+b+c+d

Paso 2. Estado de la hipótesis: La hipótesis  $H_0$  es que las especies son independientes (No hay asociación).

Paso 3. Cálculo de la prueba estadística: Para probar la asociación se calculan los valores esperados para cada celda de la tabla 2x2, si las ocurrencias de las especies A y B son de hecho independientes y se comparan con los valores observados.

Se puede usar la prueba  $\chi^2$  para probar  $H_0$  de independencia en la tabla 2x2.

Los valores esperados para cada celda son:

$$E(a) = \frac{(a+b)(a+c)}{N} = \frac{r * m}{N}$$

$$E(b) = \frac{(a+b)(b+d)}{N} = \frac{m * s}{N}$$

$$E(c) = \frac{(a+c)(c+d)}{N} = \frac{r * n}{N}$$

$$E(d) = \frac{(b+d)(c+d)}{N} = \frac{s * n}{N}$$

y la Chi cuadrado se calcula así:

$$X^2 = \frac{(a - E(a))^2}{E(a)} + \dots + \frac{(d - E(d))^2}{E(d)}$$

Y de forma más simple así:

$$X^2 = \frac{N * (ad - bc)^2}{mnr s}$$

Es de más simple uso y no requiere el cálculo de valores esperados, ni las diferencias cuadráticas entre valores observados y esperados, la significancia de esta  $X^2$  es probada con base en la  $X^2$  teórica.

La tabla de contingencia 2x2 posee un grado de libertad, mientras que una tabla de contingencia con r filas y c columnas posee (r-1)\*(c-1) grados de libertad.

El valor teórico de  $X^2$  con una significancia del 5% y un grado de libertad es igual a 3,84.

Si  $X^2$  calculado >  $X^2$  tabla, rechaza la hipótesis nula que la co-ocurrencia de las dos especies A y B es independiente, es decir, que ambas están asociadas.

Si  $X^2$  calculado <  $X^2$  tabla, acepta la hipótesis nula que la co-ocurrencia de las especies A y B es independiente, es decir, ambas especies no están asociadas.

Existen dos tipos de asociaciones:

1. Positiva: Si  $a > E(a)$ , el par de especies ocurrieron juntas más de lo esperado que si estuvieran independientes.
2. Negativa: Si  $a < E(a)$ , el par de especies ocurrieron juntas menos de lo esperado si fueran independientes.

La comparación de  $a$  (observado) con respecto a  $E(a)$  es:

$$a - E(a) = \frac{(ad - bc)}{N}$$

Sin embargo, se aplica una  $X^2$  corregida para evitar valores sesgados, con base en la fórmula de corrección de Yates:

$$X^2 = \frac{N \left[ (ad) - (bc) - \left( \frac{N}{2} \right) \right]^2}{mnrS}$$

### **5.2.1. Medida de Asociación de dos especies**

Hubalek (1982), citado por Ludwing y Reynolds (1988), identificó cinco condiciones para que la medida fuera admisible:

1. Cada índice de asociación debe dar su valor mínimo en  $a=0$ , cuando dos especies no se encontraron juntas.
2. El valor máximo del índice se da cuando ambas especies ocurren siempre juntas, cuando  $b=c=0$ .
3. El índice de asociación debe ser simétrico, es decir, que el valor del índice sea el mismo a pesar de cual especie sea designada, A o B.
4. El índice debe ser capaz de discriminar entre asociaciones positivas y negativas.
5. El índice debe ser independiente de  $d$ , que es el número de ausencias adjuntas (asociadas). Los valores de  $d$  son pequeños (limitados).

Para medir la asociación de dos especies se aplican los índices de Ochiai, Dice y Jaccard, los cuales:

- Cuando no hay asociación son iguales a cero.
- Cuando hay máxima asociación son iguales a uno.

Los índices de Ochiai y Dice son promedios de las relaciones  $a/m$  y  $a/r$  que son el número de ocurrencias adyacentes de las 2 especies comparadas a las ocurrencias totales de las especies A y B, respectivamente.

Índice de Ochiai (OI): Basado en la media geométrica de  $a/m$  y  $a/r$ , que es:

$$OI = \frac{a}{\sqrt{a+b}\sqrt{a+c}}$$

Índice de Dice (DI): Basado en la media armónica de  $a/m$  y  $a/r$ , que es:

$$DI = \frac{2a}{2a+b+c}$$

Este índice se ejecuta bien cuando  $N > 20$ .

Índice de Jaccard (JI): Este índice es la proporción del número de unidades de muestreo donde al menos una de las especies es encontrada.

$$JI = \frac{a}{a+b+c}$$

Este índice no es viable con tamaños de muestra pequeños ( $N < 10$ ).

### **5.2.2. Asociación Interespecífica (caso de múltiples especies)**

El número de todos los posibles pares de asociaciones que pueden ser calculados y se incrementa rápidamente de acuerdo a la ecuación  $s(s-1)/2$ , donde  $s$  es el número de especies.

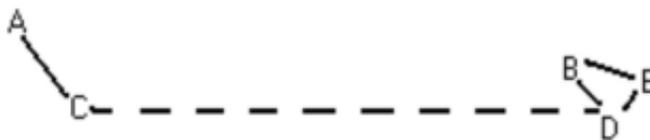
Existen dos formas de diagramar las múltiples asociaciones:

Diagrama 1: Matriz de comparación de asociación de especies, donde las posiciones en la matriz pueden ser reordenadas de tal forma para posicionar especies con valores de alto índice de significancia positiva a lo largo de la diagonal de la matriz.

ESPECIES

	A	B	C	D	
A	/				
B					I
C					
D				/	

Diagrama 2. Diagrama de constelación de especies (plexo), es una figura bidimensional de plexo, que sirve para sumar los pares de asociaciones encontrados dentro de una comunidad ecológica. Dichos diagramas de plexos son construidos de tal forma que las especies con asociaciones positivas son posicionadas muy cerca y aquellas con asociaciones negativas son posicionadas lejos:



La distancia entre especies se puede organizar en una escala de manera proporcional a la magnitud de un índice de asociación, por ejemplo, las dimensiones de las distancias entre especies se puede escalar entre 0 y 1, para reflejar los valores de los índices OI, DI y JI. Aunque estos diagramas son dibujados con posibilidad de error, mientras que el número de especies se incrementa, requiere mucho esfuerzo para aminorar distorsiones.

La técnica de pares es inadecuada cuando  $s > 2$ . Schluter (1984), citado por Ludwig y Reynolds (1988), sugirió un acercamiento usando la relación de varianza (VR), derivada de un modelo de asociación nula para probar simultáneamente asociaciones significativas. El índice de asociación VR se deriva de una tabla de presencia/ausencia, donde se obtiene el estadístico W, para probar partidas significativas desde el valor esperado de no asociación, donde W se aproxima a la distribución  $\chi^2$ , así:

Paso 1. Suma de datos a partir de la sábana:

Especie	UNIDADES DE MUESTREO (UM)						total especies
	1	2	3	4	5	... N	
1	1	0	0	0	1	n1	
2	0	0	1	0	1	n2	
3	1	1	0	0	1	n3	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
S	1	1	1	1	0	ns	
UM totales							

Paso 2. Estado de la hipótesis: La hipótesis nula se basa en que no hay asociación entre las s especies, es verdadera bajo dos condiciones:

- Que las especies son independientes.
- Las asociaciones positivas y negativas entre especies se cancelan unas con otras; además, la hipótesis alternativa se basa en que existe una asociación positiva o negativa neta.

Paso 3. Cálculo de la prueba estadística: Se calcula primero la varianza total de la muestra para la ocurrencia de s especies en la muestra así:

$$\sigma_T^2 = \sum P_i * (1 - P_i)$$

Donde  $P_i = \frac{n_i}{N}$  (refiérase a la sábana)

Luego se estima la varianza total para las especies:

$$S_T^2 = \frac{\sum_{j=1}^N (T_j - t)^2}{N}$$

Donde  $t$  es el número promedio de las especies por muestra.  
Posteriormente se calcula la relación de varianza:

$$VR = \frac{S_T^2}{\sigma_T^2}$$

La cual sirve como un índice para todas las asociaciones de especies.  
El valor esperado bajo la  $H_0$  de independencia es 1.

Independencia=1

$VR > 1$ , asociación positiva

$VR < 1$ , asociación negativa

Luego se calcula  $W$ :

$$W = VR * N$$

Según Ludwig y Reynolds (1988), cuando las especies no se asocian, existe un 90% de probabilidad que  $W$  oscile entre:

$$X_{0.05, N}^2 < W < X_{0.95, N}^2$$

La relación de varianza es apropiada para grandes colecciones de especies, aunque esta técnica también detecta asociaciones significativas donde las pruebas para pares no sirven.

Ejemplo:

Se realizó el conteo de individuos por especie para 3 especies de orquídeas en cierto corredor biológico no intervenido, observándose lo siguiente en 6 unidades de muestreo evaluadas:

	UM1	UM2	UM3	UM4	UM5	UM6	Total
Spp1	2	1	2	1	2	4	12
Spp2	3	0	0	3	2	5	13
Spp3	0	2	2	0	6	0	10
Total	5	3	4	4	10	9	36

Seguidamente se realiza la tabla de presencia/ausencia de individuos

	UM1	UM2	UM3	UM4	UM5	UM6	Total
Spp1	1	1	1	1	1	1	6
Spp2	1	0	0	1	1	1	4
Spp3	0	1	1	0	1	0	3
Total	2	2	2	2	3	2	13

Después se realiza la prueba de asociación por pares de especies:  
Especie 1 Vs Especie 2, se construye una tabla de contingencia:

		Especie 2		
		Presencia	Ausencia	
Especie 1	Presencia	a= 4	b= 2	m= 6
	Ausencia	c= 0	d= 0	n= 0
		r= 4	s= 2	

Hipótesis: La ocurrencia de las especies 1 y 2 en las 6 unidades de muestreo es independiente.

Se calcula el estadístico:

$$E(a) = \frac{4 * 6}{6} = 4 \quad (\text{Es igual al valor de a}).$$

Luego,

$$X^2 = \frac{6[(4*0) - (2*0)]^2}{(6*0*4*2)} = 0$$

Y con la corrección de continuidad (Corrección de Yates):

$$X^2 = \frac{6 \left[ (4*0) - (2*0) - \left( \frac{6}{2} \right) \right]^2}{(6*0*4*2)} = 0$$

Como  $X^2_{calculado} < X^2_{tabla} (X^2_{0.05}(1) = 3.84)$  se acepta la hipótesis que la ocurrencia de ambas especies es independiente en las unidades de muestreo.

Para evaluar la asociación de las especies 1 y 2, se calculan los índices de Dice, Jaccard y Ochiai:

Índice de Ochiai:

$$OI_{1,2} = \frac{4}{\sqrt{4*2}\sqrt{4*0}} = 0 \quad \text{Indica no asociación.}$$

Índice de Dice:

$$DI_{1,2} = \frac{2(4)}{2(4)+2+0} = 0.8$$

Es de resaltar que, este índice se ejecuta adecuadamente cuando  $N \geq 20$ .

Índice de Jaccard:

$$JI_{1,2} = \frac{2}{2+4+0} = 0.333$$

La probabilidad de que al menos una de las 2 especies sea encontrada en las diferentes unidades de muestreo es del 33.3%

**Nota:**

Cuando N (obtenido en la tabla de contingencia) es pequeño ( $N < 10$ ), no es recomendable estimar los índices de Jaccard y Ochiai.

De igual manera se realizan los cálculos para especie 1 Vs Especie 3 y la Especie 2 Vs Especie 3.

Los resultados totales obtenidos se presentan en el siguiente cuadro resumen:

Índice	1 Vs 2	1 Vs 3	2 Vs 3
OI	0	0.71	0.289
DI	0.8	0.67	0.286
Jl	0.333	0.5	0.167

No obstante, existe el análisis de asociación interespecífica para el caso de múltiples especies, así:

$H_0$  No hay asociación entre las especies.

$H_1$  Existe asociación entre las especies.

Se calcula la varianza poblacional total para las ocurrencias de las especies en la muestra con base en la tabla de presencia/ausencia:

$$\sigma_t^2 = \left(\frac{6}{6}\right) * \left(1 - \frac{6}{6}\right) + \left(\frac{4}{6}\right) * \left(1 - \frac{4}{6}\right) + \left(\frac{3}{6}\right) * \left(1 - \frac{3}{6}\right) = 0.4711$$

Seguidamente se calcula t, a partir de la misma tabla con base en los totales Tj:

$$t = \frac{2+2+2+2+3+2}{6} = 2.17$$

Después se calcula la varianza muestral total:

$$s_t^2 = \frac{1}{6} \left[ (2-2.17)^2 + (2-2.17)^2 + (2-2.17)^2 + (2-2.17)^2 + (3-2.17)^2 + (2-2.17)^2 \right] = 0.1389$$

Se calcula la relación de varianza (VR):

$$VR = \frac{0.1389}{0.4711} = 0.2948$$

Seguidamente,

$$W = 6 * 0.2948 = 1.7688$$

Se calculan,

$$X^2_{0.95}(6) = 1.64 \quad \text{y} \quad X^2_{0.05}(6) = 12.59$$

Dado que,

$$1.64 < W_{(1.7688)} < 12.59$$

Se acepta la hipótesis nula que no existe asociación entre las especies.

### 5.3. Covariación interespecífica

Esta sirve para entender si el incremento o decremento de poblaciones de especies que están juntas es de índole inverso o directo en la proporción.

El uso de coeficientes de correlación para medir la relativa intensidad de covariación en la abundancia de datos de un par de especies se denota en este capítulo.

Si los datos de abundancia para las especies está disponible (por ejemplo: porcentaje de cobertura, biomasa, densidad) se pueden generar cuestionamientos adicionales con respecto a la afinidad de las especies; es decir, como la abundancia de una especie afecta a las demás.

Dos especies pueden ser afines estrechamente, pero pueden tener una covariación negativa (correlación  $-1$ , a medida que una aumenta, la otra decrece). Entonces, la correlación se entiende en la ecología, como la covariación de tales especies.

Es por lo anteriormente expuesto que, el coeficiente de correlación de Pearson se calcula y se prueba la hipótesis que la correlación existe,

cuando entre los datos existe una distribución normal. En caso de existir una distribución no paramétrica de la serie de datos se aplica el coeficiente de correlación de Spearman.

*Coefficiente de correlación de Pearson:*

Paso 1. Hipótesis nula: Las abundancias de especies no están correlacionadas.

Paso 2. Cálculo del coeficiente de correlación de Pearson,  $\gamma$  entre las i-ésima y k-ésima especies así:

$$\gamma_{(i,k)} = \frac{\sum_{j=1}^N y_{ij} y_{kj} - \left[ \frac{\left( \sum_{j=1}^N y_{ij} \right) \left( \sum_{j=1}^N y_{kj} \right)}{N} \right]}{\sqrt{\left( \sum_{j=1}^N y_{ij}^2 - \frac{\left( \sum_{j=1}^N y_{ij} \right)^2}{N} \right) \left( \sum_{j=1}^N y_{kj}^2 - \frac{\left( \sum_{j=1}^N y_{kj} \right)^2}{N} \right)}}$$

$y_{ij}$ : Abundancia de la i-ésima especie en la j-ésima unidad de muestreo

$y_{kj}$ : Abundancia de la k-ésima especie en la j-ésima unidad de muestreo

La correlación oscila entre -1 y 1 así:

$$-1 \leq \gamma \leq 1$$

Paso 3. Prueba de la hipótesis nula:  $\gamma$  se calcula usando los datos muestrales, que es una estimación del coeficiente de correlación poblacional  $\rho$ .

Para probar la presencia de una correlación significativa entre las variables  $y_i$  e  $y_k$ , se prueba la hipótesis que  $H_0: \rho = 0$  y  $H_1: \rho \neq 0$ .

Para probar  $H_0$ , se compara el valor absoluto de  $\gamma$  ( $|\gamma|$ ) con el valor crítico dado por Rohlf y Sokal con grados de libertad iguales a N-2 y con el 5% de significancia.

Si  $|\gamma| >$  Valor crítico, entonces rechaza  $H_0$ .

$\gamma$  tiene algunas propiedades que limitan su utilidad como índice de afinidad:

1. Tiende a exagerar con base en la presencia de valores muy grandes en la serie de datos.
2. Puede producir correlaciones donde hay muchos ceros en los datos.
3.  $\gamma$  Sólo se puede aplicar cuando existe un comportamiento lineal entre las abundancias de las especies en las unidades de muestreo.

*Coefficiente de correlación de Spearman:*

Cuando los datos no se distribuyen en forma normal, es aplicado a mediciones no paramétricas.

Paso 1. se establece la hipótesis nula que las abundancias de las especies evaluadas no están correlacionadas.

Paso 2. Se calcula el coeficiente de correlación de Spearman, con base en los datos ordenados de mayor a menor para  $y_i$  e  $y_k$ , aplicando la siguiente fórmula:

$$\gamma_s = 1 - \frac{6 * \sum_{j=1}^N d_j^2}{N^3 - N}$$

Donde  $\sum_{j=1}^N d_j^2$  es la sumatoria de las diferencias al cuadrado de los

pares de datos, después de haber sido organizados.

$\gamma_s$  se halla con ambos  $y$  ranqueados, este también oscila entre -1 y 1.

Paso 3. Se prueba  $H_0$ , que las abundancias de las especies no están correlacionadas, se usa la misma prueba y razonamiento que para Pearson.

### Covariaciones interespecíficas múltiples.

Se establece como un número de especies en pares que covarían y posee dos formas de diagramación de los patrones de covariación entre muchas especies, los cuales fueron planteados anteriormente (ver Asociación Interespecífica (caso de múltiples especies)).

Ejemplos:

#### ● Pearson.

Los siguientes son los datos de biomasa fresca acumulada en kilogramos a los 180 días después de establecidas 2 especies de arbustos en 10 unidades de muestreo para cierta zona agroecológica de altura de 1550 m.s.n.m y temperatura de 21° C en promedio:

Especie 1: 2.31 2.53 2.64 2.84 2.85 2.43 2.31 2.82 2.51 2.73  
Especie 2: 1.40 1.53 1.62 1.81 1.75 1.43 1.40 1.70 1.54 1.68

$H_0$  Las especies no están correlacionadas al evaluar la variable biomasa.

$$r_{(0,2)} = \frac{(2.31*1.40) + (2.53*1.53) + \dots + (2.73*1.68) - \frac{(2.31+2.53+\dots+2.73)*(1.40+1.53+\dots+1.68)}{10}}{\sqrt{\left[ \frac{2.31^2 + 2.53^2 + \dots + 2.73^2 - \frac{(2.31+2.53+\dots+2.73)^2}{10}}{10} \right] \left[ \frac{1.40^2 + 1.53^2 + \dots + 1.68^2 - \frac{(1.40+1.53+\dots+1.68)^2}{10}}{10} \right]}}$$

$$r_{(0,2)} = 0.9819$$

Existe correlación entre ambas especies.

Sin embargo, se debe probar la hipótesis con base en la no correlación, con base en cero.

$$H_0 : r_{(0,2)} = 0$$

$$H_1 : r_{(0,2)} \neq 0$$

A través de T-student.

$$t_{calc} = \frac{0.9819 - 0}{\sqrt{\frac{1 - 0.9819^2}{10 - 2}}} = 14.66 \quad t_{tabla} \rightarrow t \frac{0.05}{2} (10 - 2) = 2.306$$

Como  $t_{calc} > t_{tabla}$  Se acepta la hipótesis alternativa ( $H_1$ ) que la correlación de ambas especies difiere significativamente de cero (No correlación).

- Spearman.

Se contaron el número de larvas parasitadas y el número de parasitoides en 6 sitios diferentes, observándose lo siguiente:

No Parasitoides (x)	No Larvas Parasitadas (y)	$x'$	$y'$	$\delta_1$	$\delta_1^2$
2	3	1	1	0	0
6	6	3	3	0	0
10	10	5	6	-1	1
8	7	4	4	0	0
4	4	2	2	0	0
12	9	6	5	1	1
					$\sum \delta_1^2 = 2$

$$\gamma_s = 1 - \frac{6 * 2}{6^3 - 6} = 0.9429$$

Existe correlación entre ambas especies.

Se realiza la prueba de hipótesis,

$$H_0 : \gamma_s = 0$$

$$H_1 : \gamma_s \neq 0$$

$$t_{calc} = \frac{0.9429 - 0}{\sqrt{\frac{1 - 0.9429^2}{6 - 2}}} = 5.66$$

$$t_{tabla} \rightarrow t_{\frac{0.05}{2}}(6 - 2) = 2.776$$

Como  $t_{calc} > t_{tabla}$  Se acepta la hipótesis alternativa ( $H_1$ ) que la correlación de ambas especies difiere significativamente de cero (No correlación).

## 6. CLASIFICACIÓN DE COMUNIDADES.

### Antecedentes.

La clasificación se define como la agrupación o formación de clusters por objetos con ciertas similitudes.

La clasificación es de suma importancia en los sistemas naturales, ya que puede ser vista como el reflejo de varios procesos que permitieron el arreglo observado de los objetos.

El primer paso para determinar una clasificación es el muestreo, ya sea este natural o artificial de datos cuantitativos o cualitativos, los cuales se agrupan según la similitud.

No obstante la clasificación puede ser jerárquica o reticulada.

*Jerárquica:* Los grupos en cualquier nivel de clasificación son subgrupos exclusivos de estos grupos en altos niveles.

*Reticulada:* los grupos son definidos separadamente y posteriormente se ordenan jerárquicamente, los cuales son encadenados en un trabajo neto.

Además, la clasificación puede ser divisiva o aglomerativa:

*Divisiva:* La colección entera de unidades de muestreo es divisiva con base en las similitudes de las unidades de muestreo para llegar a una agrupación final.

*Aglomerativa:* Como su nombre lo indica, las unidades de muestreo individuales son combinadas y recombinadas sucesivamente para formar grupos de unidades de muestreo más amplios.

También la clasificación puede ser de índole monotético y politético.

*Monotético:* La similitud de cualquier dos grupos de unidades de muestreo está basada en el valor de una sola variable.

*Politético*: La similitud de cualquier dos grupos de unidades de muestreo está basada en la medición de numerosas variables.

Al analizar la clasificación, se realiza con base en las especies o las unidades de muestreo, su relación se da en forma geométrica y se conoce como hiperespacio, que se conceptualiza como dimensional-S, que es una dimensión para cada especie en la muestra de S-especies.

La distancia entre unidades de muestreo en el espacio S representa su similitud.

## **6.1 Funciones de semejanza**

Cuando se deben llevar a cabo comparaciones de muestras de animales y/o plantas cuando se direccionan los cuestionamientos acerca de la estructura de una comunidad, las muestras pueden ser obtenidas de varias localidades en el terreno o en el mismo sitio pero en diferentes tiempos.

En sí, las funciones de semejanza se basan en la cuantificación de la similitud entre dos objetos, con base en sus observaciones, a través de un juego de descriptores.

Los objetos de interés aquí, son las unidades de muestreo o las muestras y los descriptores son mediciones de abundancia de especies entre unidades de muestreo (modo - Q). Existen dos tipos de semejanza de modo-Q en forma de funciones: coeficientes de similitud y coeficientes de distancia.

La similitud varía de 0 (Cuando un par de unidades de muestreo son completamente diferentes) a 1 (Cuando las unidades son idénticas). Para los coeficientes de distancia, su interpretación se realiza de manera inversa, aquí el valor tiende a infinito, que es el mismo comportamiento de cero.

**Coefficientes de similitud:** Son aquellos basados en la presencia/ ausencia de individuos en las unidades de muestreo, siendo los más aplicados los índices de Dice, Ochiai y Jaccard.

Coeficientes de distancia: Son tres tipos de mediciones.

- " Grupo-E: Coeficientes de distancia euclidiana.
- " Grupo-BC: Índice de similitud de Bray Curtis.
- " Grupo-RE: Medición de distancia euclidiana relativa.

Distancias del grupo E:

Distancia Euclidiana (DE): Es una ecuación para calcular la distancia entre 2 puntos  $UM_j$  y  $UM_k$  en el espacio euclidiano,

$$DE_{jk} = \sqrt{\sum_{i=1}^S (x_{ij} - x_{ik})^2}$$

DE enfatiza en las más grandes diferencias en abundancia de especies entre unidades de muestreo, mientras que cada diferencia cuadrática es sumada.

Distancia cuadrática euclidiana (DCE): Es la raíz cuadrada de DE,

$$DCE_{jk} = \sum_{i=1}^S (x_{ij} - x_{ik})^2$$

Distancia euclidiana promedio (DEP): Es similar a DE, pero la distancia final está sobre una muy pequeña escala,

$$DEP_{jk} = \sqrt{\frac{\sum_{i=1}^S (x_{ij} - x_{ik})^2}{S}}$$

Distancia absoluta (DA): Esta medida es el valor absoluto de la suma de las diferencias tomadas a través de las especies S, así:

$$DA_{jk} = \sum_{i=1}^S |x_{ij} - x_{ik}|$$

Es de menos énfasis en diferencias más grandes que las previas tres mediciones, además, las diferencias más pequeñas que son dadas son relativamente más grandes en peso en la distancia final.

Distancias del Grupo BC: Este grupo está representado por un solo índice, Bray Curtis (1957); sin embargo, el primer paso es calcular el porcentaje de similitud (PS) entre las unidades de muestreo  $j$  e  $k$ , así:

$$PS_{jk} = \frac{2 * \left( \sum_{i=1}^S [\min(x_{ij} - x_{ik})] \right)}{\left( \sum_{i=1}^S x_{ij} \right) + \left( \sum_{i=1}^S x_{ik} \right)} * 100$$

Al no multiplicar por cien, entonces PS oscila entre 0 y 1:  $0 \leq PS_{jk} \leq 1$ . Además, el PS entre la  $j$ -ésima y la  $k$ -ésima unidades de muestreo es el numerador de 2 veces la suma del mínimo del par de observaciones  $x_{ij}, x_{ik}$ . (la abundancia de especies marcada entre cada par de unidades de muestreo), dividido por el total de todas las abundancias de especies para las dos unidades de muestreo. Para cualquier par de unidades de muestreo con abundancias idénticas de especies, su similitud es completa, que es PS=100%, la distancia de disimilitud se denota PS\* (PS complemento) y es:

$$PD = 100 - PS$$

PD también se puede calcular así:

$$PD_{jk} = 1 - \frac{2 * \left( \sum_{i=1}^S [\min(x_{ij} - x_{ik})] \right)}{\left( \sum_{i=1}^S x_{ij} \right) + \left( \sum_{i=1}^S x_{ik} \right)} \quad \text{entonces } 0 \leq PS_{jk} \leq 1$$

Distancias del grupo RE: Este grupo contiene índices de distancia que son expresados en escales estandarizadas o relativas.

Distancia euclidiana relativa (DER): Esta medición incorpora los totales de abundancia de especies dentro de cada unidad de muestreo y la medición de distancia final es relativa, estandarizada, a las diferencias

en las abundancias totales de las unidades de muestreo:

$$DER_{jk} = \sqrt{\sum_{i=1}^S \left[ \left( \frac{x_{ij}}{\sum_i^S x_{ij}} \right) - \left( \frac{x_{ik}}{\sum_i^S x_{ik}} \right) \right]^2}$$

Distancia absoluta relativa (DAR): Aplica corrección a la abundancia relativa y es:

$$DAR_{jk} = \sum_{i=1}^S \left| \left( \frac{x_{ij}}{\sum_i^S x_{ij}} \right) - \left( \frac{x_{ik}}{\sum_i^S x_{ik}} \right) \right| \quad \text{entonces } 0 \leq DAR \leq 2$$

Distancia Cuerda (DCC): Da importancia en las proporciones relativas de las especies en las unidades de muestreo y menos importancia en las cantidades absolutas.

Este es dado por la proyección de las unidades de muestreo dentro de un círculo con un radio unitario a través del uso de las direcciones de los cosenos y la medición es la distancia-cuerda entre dos unidades de muestreo después de tal proyección y está dada así:

$$DCC_{jk} = \sqrt{2 * (1 - CCos_{jk})} \quad \text{donde } CCos_{jk} = \frac{\sum_{i=1}^S (x_{ij} x_{ik})}{\sqrt{(\sum_i^S x_{ij}^2) (\sum_i^S x_{ik}^2)}}$$

En el caso de evaluar presencia/ausencia, Ccos es el mismo índice de Ochiai.

Distancia geodésica (DG): Es la distancia a lo largo del arco de la unidad circular, después de la proyección de las unidades de muestreo dentro de círculos de radios unitarios:

$$DBG = Arc.Cos(CCos_{jk}) \quad \text{entonces } 0 \leq DG \leq \pi/2$$

Para sumar las distancias calculadas entre todos los posibles pares de unidades de muestreo, basadas en cualquier similitud o medidas de distancia descritas previamente, es conveniente crear una matriz  $UM \times UM$  para distancias o similitudes.

El examen de esta matriz revela rápidamente la distancia entre dos UMs de interés.

Es sobre esta matriz de distancia que operan las estrategias de clustering al clasificar comunidades.

Ejemplo:

Se están investigando 3 especies forestales al interior de un estudio de sucesión vegetal. Para tal fin, se contaron el número de éstas en 2 unidades de muestreo, obteniéndose lo siguiente:

Unidades de muestreo

Especie	(1)	(2)	(3)
(1)	7	18	0
(2)	9	0	3
(3)	14	2	7

Se calculan las diferencias, sumas y sumas de cuadrados necesarias para el cálculo de distancia entre las unidades de muestreo 1 y 2, respectivamente.

Especie	(1)	(2)	Diferencia	Diferencia <sup>2</sup>	$\sum$ Especies
(1)	7	18	-11	121	25
(2)	9	0	9	81	9
(3)	14	2	12	144	16
$\sum UMs$	30	20			
$\sum (UMs)^2$	326	328			

Seguidamente se calculan las distintas distancias:

- Distancia Euclidiana:

$$DE = \sqrt{(7-18)^2 + (9-0)^2 + (14-2)^2} = 18.6$$

- Distancia Euclidiana Cuadrática (Con base en las diferencias al cuadrado):

$$DEC = 121 + 81 + 144 = 346$$

- Distancia Euclidiana Promedio:

$$DEP = \sqrt{\frac{346}{3}} = 10.74$$

- Distancia Absoluta :

$$DA = |7-18| + |9-0| + |14-2| = 32$$

- Distancia Absoluta Promedio:

$$DAP = \frac{32}{3} = 10.7$$

- Disimilitud de Bray-Curtis (Porcentaje de disimilitud):

$$PD = 1 - \left[ \frac{2(7+0+2)}{47+6} \right] = 0.6604 = 66.04\%$$

- Distancia Euclidiana Relativa:

$$DER = \sqrt{\left[ \left( \frac{7}{30} - \frac{18}{20} \right)^2 + \left[ \left( \frac{9}{30} - \frac{0}{20} \right)^2 + \left[ \left( \frac{14}{30} - \frac{2}{20} \right)^2 \right] \right]}$$

$$DER = \sqrt{0.669} = 0.818$$

- Distancia Absoluta Relativa:

$$DAR = \left| \left( \frac{7}{30} \right) - \left( \frac{18}{20} \right) \right| + \left| \left( \frac{9}{30} \right) - \left( \frac{0}{20} \right) \right| + \left| \left( \frac{14}{30} \right) - \left( \frac{2}{20} \right) \right|$$

$$DAR = |-0.67| + 0.3 + 0.37 = 1.34$$

- Distancia-Cuerda:

$$DC = \frac{[(7*18)+(9*0)+(14*2)]}{\sqrt{326+328}} = 0.471$$

- Distancia Geodésica (En radianes):

$$DG = \text{ArcCos}(0.471) = 1.08$$

Estos cálculos se llevan a cabo de igual manera para los pares de unidades de muestreo 1-3 y 2-3, respectivamente.

Se obtiene así el siguiente cuadro de distancias:

Grupo Distancia	Medida de Distancia	Relación UM1-UM2	Relación UM1-UM3	Relación UM2-UM3
E	DE	18.6	11.6	18.9
	DEC	346	134	358
	DEP	10.7	6.7	10.9
	DA	32	20	26
	DAP	10.7	6.7	8.7
BC	PD	0.66	0.50	0.87
ER	DER	0.818	0.33	1.12
	DAR	1.34	0.47	1.80
	DC	0.471	0.43	1.34
	DG	1.08	0.43	1.47

Al observar las distancias del grupo E se aprecia que los valores más bajos están dados para la relación existente entre las unidades de muestreo 1 y 3, lo cual indica que ambas unidades son similares al analizar el número de individuos por especie encontrados en cada una de ellas. No obstante, esto puede ser corroborado por las siguientes 5 distancias de los grupos BC y ER (PD, DER, DAR, DC y DG, respectivamente), ya que para las unidades de muestreo 1 y 3 se aprecian las menores distancias.

## **6.2. Análisis de Asociación**

Un grupo se considera homogéneo cuando la asociación de especies desaparece (Ludwing y Reynolds (1988)).

Lo cual indica también que algunas especies son más sensibles que otras a los factores ambientales que controlan la estructura de la comunidad y además esas especies pueden formar la base de una clasificación.

El pionero de este criterio fue David Boodall (citado por Ludwing y Reynolds (1988)), quien afirmó: "Si la inspección de un juego de datos de cuadrantes muestra que las cantidades de dos especies están asociadas, es probable que la distribución de una o ambas de estas especies a través del área no es uniforme y que los factores del hábitat que las controlan no son uniformes".

Dada una colección de datos de unidades de muestreo, si hay especies que exhiben asociaciones significativas con otras especies (con base en  $X^2$  para las tablas 2x2), entonces la colección se considera homogénea.

La heterogeneidad puede ser reducida por la subdivisión de las unidades de muestreo, para que las asociaciones específicas que son significativas desaparezcan.

Al repetir dicha operación se reduce la heterogeneidad al interior dentro de los diferentes grupos, mientras que la heterogeneidad entre grupos se incrementa.

Es de resaltar que, el análisis de asociación es una técnica monotética, divisiva y jerárquica. El análisis de asociación clasifica unidades de muestreo, comparando pares de asociaciones de especies, las cuales se basan solamente en presencia/ausencia de los datos en las unidades de muestreo.

Para el análisis en mención, se deben llevar a cabo los siguientes pasos:

**Paso 1.** Se construyen las tablas de contingencia de la forma 2x2, para cada uno de los posibles pares de asociaciones de especies para sus datos de presencia/ausencia.

**Paso 2.** Se calcula  $X^2$  para cada tabla de contingencia y se compara con el valor crítico teórico. Si ningún par de datos es significativo,  $(S*(S-1))/2$  entonces el grupo se considera homogéneo y el análisis termina.

Si una especie tiene asociaciones significativas, la colección de unidades de muestreo se considera heterogénea y sigue el paso 3.

**Paso 3.** Se lleva a cabo la suma significativa de cuadros  $X^2$  para cada especie. Si una especie no tiene asociaciones significativas, su suma sería cero. Si hay cinco  $X^2$  significativas, entonces se suman solo estos.

**Paso 4.** Se selecciona el divisor de especies, con base en la especie con el valor  $X^2$  más alto.

**Paso 5.** Se realiza el sorteo de las unidades de muestreo, con base en la partición en dos grupos, basadas en presencia/ausencia del divisor de especies. Este resultado es una subdivisión del grupo original de las unidades de muestreo dentro de dos grupos, uno con el divisor de especies para presencia y otro para ausencia.

Para evaluar si los grupos resultantes son ahora homogéneos, un nuevo de  $X^2$  debe construirse para ambos nuevos grupos. Si ningún grupo posee asociaciones significativas, indica homogeneidad de los mismos en la colección de unidades de muestreo.

En caso de presentarse asociaciones significativas de especies, se deben repetir de nuevo todos los pasos; es decir, que el proceso sigue hasta que la homogeneidad se logre en todos los grupos de especies. Para detener el proceso se recomienda la prueba de ensayo y error, evitando la excesiva fragmentación de grupos; además, no es necesario fragmentar pocas unidades de muestreo.

Madgwick y Desrochers (1972), citados por Ludwing y Reynolds (1988), sugirieron como mínimo seis unidades de muestreo por grupo antes que la subdivisión sea considerada; además, las especies que son raras pueden ser excluidas del análisis.

Ejemplo:

Se tienen los registros obtenidos al evaluar el número de individuos para 4 especies de plantas nativas de Colombia en un corredor biológico. Fueron tomadas 8 unidades de muestreo y se encontró:

*Unidades de Muestreo*

		1	2	3	4	5	6	7	8
<i>Especies</i>	1	13	9	17	0	6	0	8	0
	2	0	7	3	9	5	5	0	2
	3	4	1	1	0	0	0	2	4
	4	5	3	3	0	2	4	0	0

Se realiza la tabla de presencia/ausencia.

*Unidades de Muestreo*

		1	2	3	4	5	6	7	8
<i>Especies</i>	1	1	1	1	0	1	0	1	0
	2	0	1	1	1	1	1	0	1
	3	1	1	1	0	0	0	1	1
	4	1	1	1	0	1	1	0	0

Seguidamente se realizan los cálculos del análisis de asociación con base en  $\chi^2$  y el tipo de asociación (A través de pruebas de hipótesis para  $\chi^2$ ).

Especie 1 vs Especie 2:

		Especie 2		Total
		Presencia	Ausencia	
Especie 1	Presencia	A 3	B 2	m 5
	Ausencia	C 3	D 0	n 3
Total		R 6	S 2	N 8

Se calcula  $\chi^2$ :

$$\chi^2 = \frac{8[(3*0)-(3*2)]^2}{5*3*6*2} = 1.6$$

$$\chi^2_{tabla} = \chi^2_{0.05}(1) = 3.84$$

Como  $\chi^2 < \chi^2_{tabla}$  se acepta la hipótesis nula, es decir, la co-ocurrencia de ambas especies no es independiente y ellas están asociadas.

Se realiza para cada par que se genere de las especies (1 vs 3, 1 vs 4, 2 vs 3, 2 vs 4 y 3 vs 4, respectivamente), obteniéndose la siguiente tabla:

Pares de Especies	Valor de $\chi^2$
1-2	1.6
1-3	1.742
1-4	1.742
2-3	1.6
2-4	0.178
3-4	0.036

Como ningún valor es mayor o igual a 3,84 ( $\chi^2$  de tabla) indica que sólo hay un grupo y existe homogeneidad en las unidades de muestreo en cuanto a las especies reportadas.

### 6.3. Análisis de Clusters

El análisis de clusters es una técnica de clasificación para posicionar entidades similares u objetos dentro de grupos conocidos como clusters.

Estos análisis sirven para localizar muestras similares dentro de los clusters, los cuales se estructuran dentro de un sistema jerárquico conocido como dendograma. Estos clusters o grupos de unidades de muestreo pueden delimitar o representar diferentes comunidades bióticas.

Cuando se tiene una serie de parcelas, transectos o unidades de muestreo y algún tipo de ensamble de unos con otros, se define la clasificación dentro de grupos llamados clusters.

El análisis de clusters requiere inicialmente el modo Q de ensamble entre las unidades de muestreo. La distancia entre los pares de combinaciones de unidades de muestreo en una colección son sumadas dentro de una distancia Unidad de muestreo x Unidad de muestreo o matriz D.

Los modelos planteados son de índole aglomerativo y comienzan con una colección de  $n$  unidades de muestreo individuales y construidas progresivamente en grupos pareados de clusters de unidades de muestreo similares.

Los pares resultantes se pueden dar por:

- Una unidad de muestreo individual con otra unidad de muestreo.
- Un individuo con un cluster existente de unidades de muestreo.
- Un cluster con otro.

El primer paso es la conformación de la matriz D por la distancia más corta entre unidades de muestreo individuales, posteriormente se construye el dendograma.

Lance y Williams (1967), citados por Ludwing y Reynolds (1988) crearon la ecuación combinatoria lineal así:

$$D_{(j,k)(h)} = \alpha_1 D(j,h) + \alpha_2 D(k,h) + \beta D(j,k)$$

Donde la distancia entre el cluster más cercano  $(j,k)$  formada por las unidades de muestreo  $j$  e  $k$  y la tercera unidad de muestreo o grupo de unidades de muestreo puede ser calculada a partir de las distancias conocidas  $D(j,k)$ ,  $D(j,h)$  y  $D(k,h)$  y los parámetros  $\alpha_1$ ,  $\alpha_2$  y  $\beta$  los cuales son los determinantes para que varíen las diferentes estrategias para determinar clusters (Ludwing y Reynolds, 1988).

Las relaciones entre las unidades de muestreo y la formación de nuevos clusters dependen de la ecuación combinatoria lineal vista con antelación:

- Dados N unidades de muestreo, hay N-1 ciclos en el análisis de clasificación.
- En el ciclo 1, se unen dos unidades de muestreo para formar un cluster.
- En el ciclo 2, otra unidad de muestreo se une al cluster inicial y así sucesivamente se forman nuevos ciclos.

Existen cuatro esquemas de peso específico, centroide (pesado y no pesado), promedio de grupo y flexible.

El análisis de cluster promedio de grupo calcula el promedio de todas las distancias entre las unidades de muestreo de diferentes grupos. Con respecto al centroide, cuando se forma un grupo es reemplazado por su media y las distancias entre clusters son las distancias entre estos promedios o centroides (Ludwing y Reynolds, 1988).

Todos los procedimientos de clusters se basan en la matriz  $D$  de todos los pares de combinaciones posibles de distancias entre unidades de muestreo, de manera secuencial:

**Paso 1.** Se obtiene el grupo inicial, con base en las dos unidades de muestreo más similares.

**Paso 2.** Se reduce la matriz  $D$ , con  $N-1$  entidades en la colección, es decir, un grupo compuesto por dos unidades de muestreo y los  $N-2$  individuos de las unidades, calculándose la ecuación combinatoria lineal, formándose una matriz  $D$ .

**Paso 3.** Se busca la matriz reducida  $D$  a partir del valor más bajo de distancia en orden para identificar entonces, el nuevo grupo a conformar.

**Paso 4.** Se repiten los pasos 2 y 3 hasta que todas las unidades de muestreo se unan en un solo grupo. El problema final en el análisis es identificar los grupos específicos o comunidades, inmediatamente después de concluido el cluster.

El dendograma se examina e interpreta por la agrupación de las unidades de muestreo, que se asocian por características que les son afines, sin embargo, se requiere del conocimiento que se tenga de la zona de estudio o de las especies evaluadas para un estudio más estructurado.

Ejemplo:

Tomando los datos del ejercicio del capítulo de análisis de asociación interespecífica para las 3 especies de orquídeas en un corredor biológico, establecer los clusters que se generan.

		Unidades de Muestreo					
		1	2	3	4	5	6
Especies	1	2	1	2	1	2	4
	2	3	0	0	3	2	5
	3	0	2	2	0	6	0

Se obtiene el grupo inicial  $D$  calculando las distancias euclidianas entre las unidades de muestreo, como se presenta aquí para las unidades 1 y 2:

$$DE = \sqrt{(2-1)^2 + (3-0)^2 + (0-2)^2 + (5-3)^2} = 3.74$$

De igual manera se hace para 1 y 3, 1 y 4 y de forma sucesiva hasta las unidades 5 y 6, obteniéndose la siguiente matriz:

		Unidades de Muestreo				
		2	3	4	5	6
$D$	1	3.74	3.61	1	6.08	2.83
	2		1	3.61	4.58	6.16
	3			3.74	4.47	5.74
	4				6.16	3.61
	5					7

Se selecciona aquel que tenga menor distancia euclidiana (en este caso las distancias 2-3 y 1-4 poseen la más baja distancia euclidiana y por ello se toma cualquiera de las 2, aunque ambas están definiendo 2 grupos).

Se tomó la distancia 2-3 para calcular las matrices reducidas  $D^i$  con base en la estrategia flexible.

Para el caso de 2-3 con respecto a 1:

$$D^i(2-3)(1) = 0.625(3.61) + 0.625(3.74) - 0.25(1) = 4.34$$

También para las demás distancias; obteniéndose la matriz:

		Unidades de Muestreo			
		2_3	4	5	6
$D'$	1	4.34	1	6.08	2.83
	2_3		4.34	5.41	7.19
	4			6.16	3.61
	5				7

Se repite el proceso, aunque ahora se ha generado otro cluster con respecto a 1-4, se halla de nuevo la matriz de distancias  $D''$  :

$$D''(1-4)(2-3) = 0.625(4.34) + 0.625(4.34) - 0.25(1) = 5.18$$

Entonces:

		Unidades de Muestreo		
		2_3	5	6
$D''$	1_4	5.18	7.4	3.8
	2_3		5.41	7.19
	5			7

Luego la matriz de distancias  $D'''$  :

$$D'''(1-4,6)(2-3) = 0.625(5.18) + 0.625(7.19) - 0.25(3.8) = 6.78$$

Y la matriz es:

		Unidades de Muestreo	
		2_3	5
$D'''$	1_4,6	6.78	8.05
	2_3		5.41

Y por último:

$$D^{\text{min}}(2-3,5)(1-4,6) = 0.625(6.78) + 0.625(8.05) - 0.25(5.41) = 7.92$$

Indica que se han formado 2 clusters claramente establecidos, el primero conformado por las unidades de muestreo 2, 3 y 5 y el segundo por las unidades 1, 4 y 6, respectivamente. Gráficamente se puede denotar así:



## BIBLIOGRAFÍA

- BURROUGH, P. 1991. Sampling designs for quantifying map unit composition. En: M.J. Mausbach y L.P. Wilding (eds), Spatial variabilities of soils and landforms, SSSA Special Publication No. 28, Madison, Wisconsin-USA. pp 89-125.
- GRAF, Esteban. Sayagués, Luis. Muestreo de la Vegetación. 2000.
- GIBBS, James P. Monitor. Numerical recipes in Pascal, Cambridge University Press.1995.
- ECOTONO, Centro para la biología de la Conservación. Primavera de 1996.
- LUDWING, Jhon A. REYNOLDS, James F. Statistical Ecology. 1988.
- MENDENHALL, William. SINCICH, Terry. Probabilidad y Estadística para Ingeniería y Ciencias. 1997.
- MONTGOMERY, Douglas. Diseño y Análisis de Experimentos. 1991.
- OVALLES, F. A. 1991. Evaluación de la variabilidad de suelos a nivel de parcela para el establecimiento de lotes experimentales en el estado cojedes. Agr. Trop. 41(1-2): 5-21.
- OVALLES, F.A. 1992. Evaluacion de la variabilidad de suelos a nivel de parcela, para el establecimiento de lotes experimentales en el Estado Cojedes. Agr. Trop. 41 (1-2): 5-21.
- SIZE, W. 1987. Use of representative samples and sampling plans in describing geologic variability and trends. En: W. Size (ed.), Use and abuse of statistical methods in the earth sciences, Oxford Univ. Press, N.Y., USA. pp 3-20.

Sitios en INTERNET donde se puede adquirir información en el área en cuestión:

BIONATUR LINKS:

CENTER FOR CONSERVATION BIOLOGY ACADEMICS:

ECOLOGÍA-COLOMBIA:

ECOLOGÍA-MÉXICO:

ECOLOGÍA-PERÚ:

GUÍAS DE MEDIO AMBIENTE:

STATISTICAL TRINING WEBSITES:

THE WORLD WIDE WEB VIRTUAL LIBRARY STATISTICS:



Este libro se imprimió en los Talleres Litográficos de OPTIGRAF S.A.,  
con un tiraje de 100 ejemplares.

La composición tipográfica se realizó  
empleando las familias Futura, Arial, Wingdings, Times New Roman  
2009



