

Dispersión Espacial: El Prerrequisito Esencial para el Muestreo

(Spatial Dispersion: The Essential Prerequisite for Sampling)

Badii, M.H., A. Guillen, E. Cerna & J. Landeros*

Abstract. The relevance of spatial dispersion as an essential prerequisite for sampling is described. Different types of spatial dispersions are defined. Equations for estimating dispersion of Poisson, uniform and clumped types are provided. The existence and the estimation of common K is noted and assessed. Furthermore, the methods of estimation of following dispersion models and indices are given: Variance-mean ratio, Alexis index, Charlier index, models of David & Moore, Morisita, Taylor, Green, Cole, Lefkovitch, Lloyed, Iwao, and Pimbert & Jarry. Examples to use each of the above mentioned models and indices are provided. Criteria for choosing the best or perfect index or model are noted. Finally, the normal distribution and the necessity to transform data with different kinds of transformations are explained.

Keywords. Dispersion, indices, models, sampling.

Resumen. Se describe la relevancia de la dispersión espacial como un requisito fundamental para el muestreo estadístico. Se definen y explican los diferentes tipos de dispersión espacial. Se proveen las ecuaciones para estimar la dispersión de tipo Poisson, uniforme, y agregada. Se nota y calcula la existencia de $K_{\text{común}}$. Además se presentan la manera de cuantificar el tipo de dispersión basad en varios modelo de uso común, Como por ejemplo: razón varianza-media, índice de Lexis, índice de Charlier, índice de David & Moore, modelo de Morisita, modelo de poder de Taylor, Modelo de Green, modelo de Cole, índice de Lefkovitch, índice de Lloyd, Modelo de Iwao, índice de Pimbert & Jarry Se presentan ejemplos para el cálculo de dispersión según cada uno de los modelos e índices anteriormente mencionados. Se presentan los criterios para seleccionar un índice perfecto. Finalmente, se discute la distribución normal y la necesidad de realizar y estimar el tipo de la transformación apropiada.

Palabras claves. Dispersión, índices, modelos, muestreo.

Introducción

La distribución espacial es una de las propiedades más características de las especies, porque produce parámetros que las segregan y estos son expresiones poblacionales del comportamiento a nivel individual. Se le puede definir como el producto de la heterogeneidad ambiental y el crecimiento de la población y reproducción, actuando sobre procesos aleatorios y dirigidos de movimiento y mortalidad (Badii et al., 1994a, 1994b, Badii et al., 2000). Los aspectos fundamentales de muestreo estadístico han sido estudiados (Badii et al., 2004, 2006; 2007a, b, c; 2010a, b, c; Badii & Catillo, 2009, 2010; Badii & Guillen 2010). La disponibilidad de microhábitats diversos y la riqueza microclimática derivada de la interacción entre la topografía y el grado de complejidad estructural de la vegetación, forman parte importante de la heterogeneidad ambiental. La contraparte está integrada por la gran plasticidad genética de mayoría de las poblaciones, las cuales al

enfrentarse al medio, se dispersan localmente. En ese punto es que ocurre la selección y desde el inicio, transcurso y final de una generación, el esquema que adopta la población, al acomodarse los individuos a las condiciones ambientales “óptimas” disponibles, es lo que se interpreta como distribución espacial. Ahora bien, se dice que este proceso de ensamble al medio es un fenómeno aleatorio.

Vamos a considerar el concepto de la distribución espacial desde el punto de vista de la mayoría de las poblaciones de este planeta, es decir, los que pertenecen a los invertebrados. Los mecanismos de orientación que usan algunas poblaciones, como por ejemplo, la mayoría de los invertebrados son el resultado de la capacidad de sus sensores ambientales, y esta capacidad, así como la propia manifestación etológica de los individuos, se comportan como variables, es decir, no exhiben la misma intensidad en todos los individuos que forman la población. Habrá algunos individuos resistentes a la desecación, pero el grueso de la población se concentrará en aquellos puntos del espacio, donde existe la humedad óptima para esa especie; pero estos puntos son de ocurrencia errática en el medio. Así mismo, el perfil higrico de los insectos que arriben a esos puntos, también será impredecible. De esta forma, se va a tener una variación en el número de individuos por punto en el espacio, llámese éste planeta, escondrijo, hoja, tallo, metro cuadrado de suelo, etc., y dicha variación esquematizada mediante una tabla de frecuencias, cuya distribución sea similar a una de las probabilísticas teóricas conocidas, es lo que se interpreta como distribución espacial.

Índices de distribución espacial: una alternativa práctica

La mayoría de las poblaciones naturales son estacionales y discontinuas, y la estabilidad solo se presenta en los modelos teóricos. En base a esto, es difícil muestrear repetidamente una población para tener una distribución de frecuencias bien definida. Los muestreos de campo raramente producen distribuciones consistentes porque son afectados erráticamente por factores como depredación, parasitismo, pérdida de hospedero, mortalidad física por lluvias, deshidratación, etc. En la práctica lo que puede medirse en un programa de muestreo es la media y la varianza, y estas dos se pueden combinar de varias formas para producir coeficientes o índices de agregación, como apoyo conceptual para el manejo de datos.

Relevancia de la distribución espacial

Cualquier investigador serio, sabe que al realizar un experimento en el laboratorio, uno de los puntos que siempre debe tener presente, es el de aplicar un buen número de repeticiones para sus tratamientos, con el objetivo de disminuir el error y conseguir una confiabilidad aceptable en sus resultados. Pero cuando el experimento es en el campo e implica estimar la media poblacional en cada parcela útil, lo más adecuado es manejar bloques; es decir el trabajo típico de un campo experimental, donde las parcelas son pequeñas y es muy posible que se esté trabajando con la misma densidad poblacional. En tales condiciones y con un coeficiente de variación pequeño (30% por ejemplo), cualquier

diferencia la consideramos efecto de los tratamientos. El problema se genera cuando el experimento abarca varias generaciones de la especie, con parcelas grandes (comerciales) e incluyendo un rango bastante amplio en la media poblacional. Aquí los resultados pueden ser seriamente defectuosos, y hasta contradictorios; algo muy rutinario en los programas de control de plagas.

Al aumentar la población de una especie, puede variar también y muy fácilmente, el patrón de distribución espacial. Si este cambio ocurre en el transcurso del experimento, se alteran los estadísticos de la muestra, y por consecuencia su aplicación será totalmente errónea. Si el experimento consiste en afinar un modelo de días-grado, para pronóstico de emergencia de los adultos, el cual será la pauta para la primera aplicación del método de control, estos resultados solo traerán consigo efectos negativos.

Por otro lado, cuando se emplea un modelo matemático para describir datos observados obtenidos de dos o más poblaciones o (por ejemplo: adultos, jóvenes, ancianos, mujeres, hombres, etc.), siempre queda la duda de si las poblaciones deben analizarse como una sola o en forma separada. Estadísticamente, cuando un cambio de la distribución de una población afecta la distribución de otra, se dice que existe dependencia entre ambas poblaciones. Lo anterior significa que cuando se incrementa el número de los adolescentes en una muestra también lo hace el número de adultos en la misma muestra. Teóricamente, puede haber una relación inversa y esto también es dependencia, tal y como ocurre en una relación parasito-hospedero. Si dos poblaciones son dependientes, un cambio en cualquiera de ellas afectaría la distribución de su suma. Así un modelo matemático particular puede describir las poblaciones en un caso, pero no en otros. En tales casos, lo que procede es usar una prueba de bondad de ajuste como la de X^2 para detectar la posible independencia poblacional entre dos fases metamórficas de una especie. Una vez que la independencia se ha confirmado, ya es posible proceder al ajuste de una distribución teórica a los datos observados, donde las muestras se basan en la suma de fases metamórficas.

Tipos de distribución espacial

El patrón de distribución espacial en término general se clasifica en uno de tres tipos básicos: aleatorio, uniforme y en agregados.

1. Patrón espacial aleatorio

Para describir este tipo de dispersión se puede imaginar un universo bidimensional cuya superficie está compuesta de muchos puntos, en este universo se refiere a una dispersión espacial aleatoria cuando: **a.** cada individuo tiene la misma probabilidad de ocupar cualquier punto o Unidad Muestral (UM), **b.** cada punto (UM) tiene la misma probabilidad de contener cualquier individuo y **c.** la presencia de un individuo en un punto (UM) es independiente de otros individuos. Cuando se reúnen estos tres rasgos, se refiere a la dispersión tipo aleatorio. Debemos resaltar que hay dos tipos de dispersión aleatoria:

- a.** La distribución normal, para los conteos altos (tamaños de muestras altas), y cuando existe una homogeneidad de

varianzas.

- b. La distribución Poisson, una indicación de rareza y cuando la varianza muestral (v) es igual a la media muestral (m). Se puede describir el modelo de Poisson mediante sólo un parámetro, ya que según este modelo, $m = v$. Se usa el modelo de bondad de ajuste (X^2) para determinar la concordancia entre los datos observados (campo) y esperados (generados por el modelo). Los datos (frecuencias) esperados se estiman mediante la ecuación: $f_{e(x)} = P_x(Sf_0)$, donde, $f_{e(x)}$ = la frecuencia esperada de la clase x , Sf_0 = la suma de las frecuencias observadas del muestreo y $P_x = (\exp^{-m})[(m^x)/(x!)]$, donde, P_x es la probabilidad de ocurrencia del individuo de la clase x , m es la media muestral ($m = S(f_0x)/Sf_0$), “!” indica factorial, y \exp es la base de logaritmo natural. El tamaño óptimo de la muestra del modelo de Poisson se calcula por: $n = (1/m)/D^2$, donde n = número de unidades muestrales, m = la media y D es el error estándar de la media como una fracción de la media. Este patrón describe una distribución espacial en la cual teóricamente, cada individuo tiene la misma probabilidad de ocurrir en cualquier área del hábitat, y su presencia es independiente en la ubicación de cualquier otro organismo. Esta condición fue escrita por Poisson en el Siglo 19. La distribución Poisson puede ser descrita por un solo estadístico, m , ya que teóricamente la diferencia entre la varianza, v y la media es igual a cero. Bajo condiciones naturales, no existe una independencia total entre los organismos, pero algunas poblaciones exhiben una distribución indistinguible de la real aleatoriedad.

Ejemplo del modelo de Poisson

Vamos a suponer que deseamos estimar el patrón de dispersión espacial de los trabajadores en las oficinas de 10 empresas (Tabla 1).

Tabla1. Número de personas por oficina por cada empresa.

Empresas	Oficinas									
	1	2	3	4	5	6	7	8	9	10
1	3	1	1	3	1	3	2	4	1	2
2	4	3	1	2	2	0	3	2	2	2
3	2	3	1	2	2	5	1	3	2	4
4	3	3	1	1	2	2	4	4	2	0
5	2	1	1	1	2	1	4	4	0	3
6	2	2	3	1	1	2	2	2	1	0
7	2	2	1	1	4	1	1	0	1	1
8	1	0	0	4	1	1	3	2	1	3
9	3	4	3	0	2	1	4	4	3	1
10	1	5	3	3	2	0	0	0	1	2

Se seleccionan 10 oficinas de forma aleatoria de cada uno de las 10 empresas. Luego se contabiliza el número de las personas en cada oficina. Por tanto, en la tabla siguiente (Tabla 1), los datos de cada celda representan el número de los trabajadores, las columnas son las oficinas y las hileras representan las empresas.

El primer paso es construir una tabla (Tabla 2) en donde se ordenan las frecuencias observadas (f_o) de cada clase (x_i) a partir de los datos de la Tabla 1. Luego en base a los datos de la Tabla 2, se calculan las probabilidades (P_x) de la ocurrencia de cada clase según el modelo de Poisson, las frecuencias observadas ($f_{e(x)}$) y los valores de X^2 .

Tabla 2. Construcción de las variables para el modelo de Poisson.

X_i	F_o	$f_o x_i$	P_x	$f_{e(x)}$	X^2
0	11	0	0.1408	14.08	0.6737
1	30	30	0.2760	27.69	0.2086
2	27	54	0.2706	27.06	0.0001
3	18	54	0.1767	17.67	0.0061
4	12	48	0.0866	08.66	1.2881
5	2	10	0.0339	03.39	0.5699
S	100	196	0.9846	98.50	2.7456

Las hipótesis nula (H_o) y la alterna (H_a) se expresan de forma siguiente:

$$H_o: f_o = f_e$$

$$H_a: f_o \neq f_e$$

Donde, f_o = Frecuencia observada, y f_e = frecuencias generadas por el modelo de Poisson. Se calculan la media (m), las probabilidades de ocurrencia de las clases (P_x), y las frecuencias esperadas de cada clase ($f_{e(x)}$) según las siguientes ecuaciones:

$$m = S(f_o x) / S f_o m = 196 / 100 = 1.96$$

$$P_x = (\exp^{-m}) [(m^x) / (x!)]$$

$$P_1 = (\exp^{-1.96}) [(1.96^1) / (1!)] = 0.1408$$

$$P_2 = (\exp^{-1.96}) [(1.96^2) / (2!)] = 0.2760$$

.

.

$$P_5 = (\exp^{-1.96}) [(1.96^5) / (5!)] = 0.0339$$

$$f_{e(x)} = P_x(S f_o)$$

$$f_{e(1)} = 0.1408 (100) = 14.08$$

$$f_{e(2)} = 0.2760 (100) = 27.60$$

.

.

$$f_{e(5)} = 0.0339 (100) = 3.39$$

Finalmente, el valor calculada de X^2 es igual a 2.7456 (Tabla 2), lo cual es menor del valor de la tabla de X^2 (9.488) a nivel de $\alpha = 0.05$ y con 4 grados de libertad. Por tanto, se apoya la hipótesis nula (H_0), en otras palabras, el patrón de distribución obedece al modelo de Poisson, es decir, el modelo de Poisson describe la dispersión de estos datos.

2. Patrón espacial uniforme

Cuando se encuentra a un individuo en un punto (UM) se reduce la probabilidad de encontrar otro individuo en el mismo UM. Este tipo de dispersión es indicativo de la competencia y territorialidad. Los datos (frecuencias) esperados se estiman mediante la ecuación: $f_{e(x)} = P_x (Sf_0)$ donde, $f_{e(x)}$ = la frecuencia esperada de la clase x , Sf_0 = la suma de las frecuencias observadas del muestreo y $P_x = \{(k!)/[x!(k-x)!]\} q^{(k-x)} p^x$, donde, P_x = la probabilidad de la ocurrencia del individuo de la clase x , K = el máximo número de los individuos por UM, x = número de clase, p es la probabilidad de la ocurrencia de un individuo en una unidad muestral y q es la probabilidad de ausencia de un individuo en una unidad muestral y “!” = factorial. El tamaño óptimo de muestra para este modelo se calcula por: $n = [(1/m) - (1/k)] / D^2$, donde, todas las notaciones como antes descritas.

Una característica de las poblaciones uniformemente distribuidas es que la varianza es menor que la media, porque hay relativamente pocas áreas de hacinamiento o densidades muy bajas en comparación a un patrón aleatorio. El espacio uniforme o regular entre los individuos, puede indicar un comportamiento competitivo o agresivo, como la territorialidad existente en algunas plantas, aves y mamíferos. Tal comportamiento es extremadamente raro en muchos invertebrados.

Ejemplo de modelo de uniforme

Se contabilizaron el número de las larvas de una mosca acuática por cada cuadrante en base a 20 cuadrantes puestos de forma aleatoria en el lecho de un arroyo (Tabla 3).

Tabla 3. Construcción de las variables para el modelo de uniforme.

X_i	f_0	$f_0 x_i$	P_x	$f_{e(x)}$	X^2
0	0	0	0.0510	1.02	1.02
1	0	0	0.0510	1.02	1.02
2	5	10	0.2100	4.20	0.14
3	10	30	0.4200	8.40	0.29
4	5	20	0.3150	6.30	0.28
S	20	60	1.0420	20.94	1.73

Se calculan los siguientes parámetros:

La media, $m = Sf_0 x_i / Sf_0 = 60/20 = 3$

La varianza $v = \sum f_0 x_i^2 - (\sum f_0 x_i)^2 / (\sum f_0 - 1) = 10 / 19 = 0.5263$

La $K = X$ máxima = 4

$p = m/K = 3/4 = 0.75$

$q = 1 - p = 0.25$, y a partir de estos valores se construyen los demás columnas de la Tabla 3.

Las hipótesis nula (H_0) y la alterna (H_a) se expresan de forma siguiente:

$$H_0: f_0 = f_e$$

$$H_a: f_0 \neq f_e$$

Donde, f_0 = Frecuencia observada, y f_e = frecuencias generadas por el modelo de uniforme.

Ahora bien, el valor calculado de X^2 es igual a 1.73 (Tabla 3), lo cual es menor que el valor tabulado (5.99) con el grado de libertad igual al número de las clases menos 3, es decir, $5 - 3 = 2$. Por tanto, se apoya a la hipótesis nula (H_0), en otras palabras, el patrón de distribución de los datos se ajusta al modelo de uniforme.

3. Patrón espacial de tipo agregado

Encontrando un individuo en un punto (UM) incrementa la probabilidad de encontrar otro individuo en el mismo UM. Este tipo de dispersión es una indicación de atracción entre los individuos; según Taylor (1961, 1984), y Taylor et al., 1979, este tipo de dispersión es la forma que se encuentra más comúnmente en la naturaleza, y basándose en la revisión de la literatura por el mismo autor, casi del 96% de los artrópodos (el grupo más diverso de todos los organismos del planeta) tienen este tipo de dispersión. Existen varios modelos que se pueden usar para describir este tipo de dispersión, sin embargo, el modelo de binomial negativa (Bliss & Fisher, 1953), ofrece una explicación más adecuada, precisa y general de la dispersión agregada. Se usa aquí también el modelo de bondad de ajuste (X^2) para determinar el ajuste entre los datos del campo y los datos esperados que se generan por el modelo, según las siguientes ecuaciones: $f_0 = n/q^k$ y $f_x = f_{(x-1)} [(m/(m+K))[(x+K-1)/x]]$, donde, f_0 es la frecuencia esperada de la clase cero, f_x es la frecuencia esperada de la clase x , $q = 1 + p$, $p = m/K$, m = la media muestral, k es el parámetro de dispersión del modelo de binomial negativa y las demás notaciones como antes descritas. Se debe distinguir entre el parámetro de K del modelo de binomial negativa y el K -valor lo cual es una estimación lineal de la fuerza de mortalidad de los factores reguladores (depredadores, parasitoides, microorganismos causantes de las enfermedades) de las poblaciones de las plagas en el contexto de control biológico (Badii & Castillo, 2008). El tamaño óptimo de la muestra para el modelo de binomial negativa es $n = [(1/m) + (1/K)]/D^2$, donde, todos los parámetros como antes descritos. Hay tres métodos para el cálculo de K de binomial negativa.

1. $K = m^2/(v-m)$
2. $\log(n/n_0) = K \log(1 + (m/K))$
3. $S(A_x/(X+K)) = n \ln(1 + (m/K))$

Donde, \log es logaritmo decimal, A_x = frecuencia acumulada, x es la clase, n es el número total de unidades muestrales, y n_0 es el número de unidades muestrales vacías, es decir sin individuos. Ahora bien, el primer método es una aproximación y los dos métodos restantes son más precisos de forma progresiva. Existen algunas características que se deben reunir para poder usar cualquiera de estos métodos (Bliss & Fisher, 1953).

En resumen, el caso usual en las poblaciones naturales es cuando la varianza es mayor que la media. En estos casos la distribución se llama “sobre-dispersa” o “de contagio”. El sobre-dispersión implica que la presencia de uno o más individuos induce la ocurrencia de otros individuos en la misma muestra. La agregación describe una condición espacial en donde la densidad esta localmente centrada. La palabra “contagio” aquí significa influenciar la distribución de otros. Ahora bien, muchos organismos de distintas especies se agregan o tienden a agregarse en respuesta a una o más condiciones bióticas o abióticas, o también por factores etológicos. Algunas de estas condiciones incluyen:

- a. Altas densidades de inmaduros como resultado de una oviposición en masas.
- b. Respuesta a microclimas tales como temperatura, humedad y luz óptima, viento y factores edáficos, etc.
- c. Respuesta a fuentes alimenticias muy espaciadas, o preferencia por ciertos biotipos de alimento, en el caso de carnívoros.
- d. Especies sociales o subsociales.
- e. Respuesta a feromonas y comportamiento copulatorio; atracción por sonido, lugares diapáusicos, etc.
- f. Parasitismo diferencial.

Independientemente de las causas de la agregación, su ocurrencia conduce a dificultades tanto en el muestreo como en el análisis. Generalmente, las densidades de las poblaciones con distribución de tipo agregada tienden a subestimarse, porque un número desproporcionadamente grande de individuos se presentan en pocos agregados, los cuales son raramente incluidos en proporciones significativas en las muestras. Para el mejor entendimiento del modelo binomial negativa, de nuevo enfatizaremos los siguientes puntos.

Modelo de binomial negativa

Para determinar la frecuencia teórica de los diferentes modelos que existen, quizás el más difícil de definir inicialmente sea la binomial negativa, porque no existe un método universalmente aceptado para determinar por primera vez el parámetro K .

Calculo de K simple

Existen tres métodos para su cálculo, los primeros dos aproximados y el tercero más preciso:

Método 1

$$K = m^2 / (v - m)$$

$$V = \sum(f_o X^2) - (\sum f_o X)^2 / n / (n-1)$$

$$m = \sum f_o X / n$$

Donde:

m = La media de la muestra.

V = La varianza muestral.

X = Clase de individuos

f_o = La frecuencia observada para la clase X .

n = Número de las unidades muestrales = $\sum f_o$.

El método es confiable cuando las poblaciones muestran un grado moderado de contagio ($K \leq 3$) y medias bajas ($m \leq 1$).

Método 2

$$\text{Log}(n/n_0) = K \log(1 + m/Kk)$$

Donde:

n = Número total de unidades de la muestra = $\sum f_o$.

n_0 = Número de las unidades muestrales con cero organismos.

K se encuentra por el método de prueba y error, empezando con el valor de la K calculada del método 1. Este método es confiable cuando 1/3 de las muestras están vacías y $m < 10$. Con medias mayores, debe haber más muestras con cero elementos (unidades muestrales vacías).

Método 3

$$\sum(A_x / (K + x)) = n \ln(1 + (m/K))$$

Donde:

x = Clase o grupo de individuos o elementos.

n = Número total de unidades de la muestra.

\ln = Logaritmo natural.

A_x = Suma de todas las frecuencias de las unidades de la muestra conteniendo más de x individuos, por ejemplo, $A_6 = \sum f_7, f_8, \dots, f_n$.

Se soluciona esta ecuación por prueba y error también. Se inicia con un valor encontrado con los dos primeros métodos. Si el lado izquierdo es grande, el valor de K debe ser menor; si el lado derecho es el mayor, el estimador de K debe elevarse.

Ajuste a la binomial negativa

Después de calcular el valor de K , se calculan las frecuencias esperadas de unidades de la muestra con x individuos con la siguiente ecuación:

$$f_{e_x} = (f_{e_{(x-1)}}) (m/(m+K)) ((x + K - 1)/x)$$

$$f_o = n / q^k$$

$$p = m/K$$

$$q = 1 + p$$

Donde,

- f_o = La probabilidad o la frecuencia esperada de unidades muestrales vacías.
- P = La probabilidad de éxito o la ocurrencia de un evento en una distribución binomial.
- q = La probabilidad de fracaso o la no ocurrencia del evento en una distribución binomial.
- f_{e_x} = Frecuencias esperadas para la clase x .
- $f_{e_{(x-1)}}$ = Frecuencias esperadas para la clase $(x-1)$.
- X = Número de clase o individuos o elementos.

Se llena la siguiente Tabla (4):

Tabla 4. Ajuste del modelo de distribución binomial negativa.

X	f_{o_x}	$f_{o_x}(x)$	$f_{o_x}(x)^2$	A_x	f_{e_x}	$X^2 = (f_{o_x} - f_{e_x})^2 / f_{e_x}$
"	"	"	"	"	"	"
"	"	"	"	"	"	"
"	"	"	"	"	"	"

Posteriormente se obtiene el valor de X^2 con: $n - 3$ grados de libertad.

Una vez obtenido el valor de K (con sus aclaraciones) el tamaño óptimo de muestra se puede calcular con la ecuación:

$$n = (k + m) / D^2 K m = (1/m + 1/k) / D^2$$

Donde:

$$D = EE/m$$

EE = El error estándar de la media

m = la media de la muestra

Ejemplo de distribución agregada

Para demostrar el patrón de dispersión agregada sería mejor utilizar el ejemplo clásico de los datos que Philip Garman obtuvo de una especie de ácaro (*Panonychus ulmi*) denominado el ácaro rojo europeo que ocasiona millones de dólares de daño a los árboles frutales en el hemisferio norte del planeta. En el campus de la Universidad de Kentucky, el Dr. Garman colecto de manera aleatoria 150 hojas del árbol del manzano y contabilizo el número de éstos ácaros por cada hoja (Tabla 5).

Tabla 5. Las clases (x_i) y las frecuencias observadas (f_o) de los datos de ácaros.

X_i	0	1	2	3	4	5	6	7
f_o	70	38	17	10	9	3	2	1

Dr. Ronald Fisher (el padre de la estadística) utilizo los datos de Dr. Garman para determinar el tipo de distribución de estos datos (Tabla 6). La media (m), la varianza (V) y el parámetro del modelo de binomial negativa (K) se calculan vía las siguientes ecuaciones:

$$m = \sum(f_o X) / \sum f_o = 1.1466$$

$$V = \sum(f_o X^2) - (\sum f_o X)^2 / \sum f_o = 2.2736$$

$$K = m^2 / (v - m) = 1.1665$$

Para los fines prácticos se utiliza el valor de K generado por el primer método de cálculo de K . Sin embargo, si uno quiere mayor precisión se puede calcular la K por el método 2 según: $\log(n/n_0) = K \log(1 + m/k)$, lo cual produce un valor de K igual a 1.000. También según las condiciones, se puede calcular este parámetro por el método 3 vía: $\sum(A_x / (K + x)) = n \ln(1 + (m/K))$, lo cual genera un valor de K igual a 1.024. El lector puede hacer su propio cálculo para producir los valores de K según los métodos 2 y 3. En este ejercicio se usa el valor de K (1.1665) generado por el método 1. El valor calculado de X^2 es igual a 2.584 (Tabla 6), lo cual es menor que el valor tabulado de X^2 (7.815) con $\alpha = 0.05$ y el grado de libertad igual a $6 - 3 = 3$. Por tanto, los datos obedecen al patrón de distribución agregado descrito por el modelo de binomial negativa. Hay que notar que las hipótesis nula (H_0) y la alterna (H_a) se expresan de forma siguiente:

$$H_0: f_o = f_e$$

$$H_a: f_o \neq f_e$$

Donde, f_o = Frecuencia observada, y f_e = frecuencias generadas por el modelo de binomial negativa.

Por tanto, se apoya la hipótesis nula que es un ajuste al modelo de binomial negativa. En otras palabras, el patrón de distribución espacial de los datos es de tipo agregado.

Tabla 6. Construcción de las variables para el modelo de binomial negativa.

X_i	f_o	$f_{e(x)}$	X^2
0	70	67.49	0.004
1	38	39.03	0.004
2	17	20.95	0.478
3	10	10.96	0.046
4	9	5.66	1.925
5	3	2.90	0.027*
6	2	1.47	
7	1	0.75	
	5 = suma de 3, 2, 1.	5.12 = suma de 2.90, 1.47, 0.75.	
Σ	150	149.21	2.584

*: Todos los f_o y $f_{e(x)}$ menor de 5 se agrupan en 1, por tanto, las 8 clases se convierten en 6 clases.

Cálculo de K común

Este método fue propuesto por Anscombe (1950) y Bliss & Owen (1958). Las formulas requeridas son:

$$\begin{aligned}
 X_i' &= (m_i)^2 / (V_i - N) \\
 Y_i' &= V_i - m_i \\
 K_i' &= X_i' / Y_i' \\
 K_{cc} &= \sum X_i' / \sum Y_i' \\
 A &= [0.5 (N - 1) K_{cc}^4] / [K_{cc} (K_{cc} + 1) - \{2 K_{cc} - 1\} / N - 3 / N^2] \\
 W_i &= A / [m_i^2 (m_i + K_i')^2] \\
 K_{cp} &= \sum W_i (X_i')^2 / \sum (W_i X_i' Y_i)
 \end{aligned}$$

Donde,

N = Número total de muestras por campo.

K_{cc} = K común crudo o K común general.

K_{cp} = K común ponderado, y los demás anotaciones están descritas e explicadas como antes mencionadas.

El modelo de binomial negativa está escrita por dos parámetros; la media (m) el exponente K . La K es un índice del grado de agregación: valores de $K \geq 8$ indican que la distribución se aproxima al patrón Poisson, y valores fraccionales se ubican la distribución en la logaritmo normal.

Aquí el resultado es un estimador ponderado del parámetro, el cual debe ser representativo de todos los campos muestreados, y para probar esto, se grafica el cociente " $Y_i' / X_i' = 1 / K_i'$ " contra el " $\log m$ " y se hace un análisis de correlación lineal. Los valores de $1 / K_i'$ no deben estar relacionados con la media y su comportamiento en una grafica como variable dependiente debe mostrar una recta paralela al eje x . En otras palabras el

coeficiente de correlación y/o regresión para esta grafica debe ser estadísticamente igual a cero para que verifique la existencia de un K común.

El parámetro K es virtualmente inútil como indicador para datos reales, ni alguna utilidad se le ha demostrado experimentalmente, si exhibe una relación densodependencia. Desafortunadamente, se ha observado que esta relación es típica para datos en donde con la ley de poder de Taylor se encuentra que $1 < b < 2$ (Taylor, 1961). En este sentido, se ha mencionado que el uso de K como una medida de agregación será descontinuado en un futuro cercano.

Ejemplo de ajuste de K común

Supongamos que los datos siguientes (Tabla 7) presentan los valores de la media y la varianza para cada uno de las seis diferentes zonas. La idea es si se puede buscar un valor de K que es común para todas estas zonas.

Tabla 7. Los valores de la media (m) y la varianza (v) para cada zona (i).

Zona (i)	Media (m)	Varianza (v)
1	75.75	539.07
2	84.00	627.14
3	42.25	257.93
4	32.50	77.43
5	48.00	172.57
6	48.62	91.12

Pasos para el ajuste de K común

1. Cálculo de K común crudo (K_{cc})

En base a las ecuaciones siguiente, se calculan los valores de X_i' , Y_i' , K_i' , W_i' , $W_i' X_i'^2$, $W_i' X_i' Y_i'$, $W_i' Y_i'^2$ (Tabla 8).

Tabla 8. Los parámetros para el cálculo de K común.

I	X^*	Y^*	K^*	W^*	$W^* X'^2$	$W^* X' Y'$	$W^* Y'^2$
1	5648.21	463.32	12.190	0.0000104	331.784	27.216	2.233
2	6951.47	543.14	12.798	0.0000070	339.227	26.505	2.071
3	1742.07	215.68	8.077	0.0001026	311.523	38.569	4.775
4	1043.34	44.93	23.221	0.0001416	154.053	6.807	0.286
5	2275.23	124.57	18.264	0.0000459	237.454	13.001	0.712
6	2348.71	42.50	55.263	0.0000182	100.344	1.816	0.033
Σ					1474..385	113.9123	10.1085

$$X_i' = (m_i)^2 / (V_i - N)$$

$$Y_i' = V_i - m_i$$

$$K_i' = X_i' / Y_i'$$

$$K_{cc} = \Sigma X_i' / \Sigma Y_i'$$

$$A = [0.5 (N - 1)K_{cc}^4] / [K_{cc} (K_{cc} + 1) - \{(2 K_{cc} - 1) / N\} - 3 / N^2]$$

$$W_i = A / [m_i^2 (m_i + K_i')^2]$$

$$K_{cp} = \sum W_i (X_i)^2 / \sum (W_i X_i Y_i)$$

$$A = [0.5 (N - 1)K_{cc}^4] / [K_{cc} (K_{cc} + 1) - \{(2 K_{cc} - 1) / N\} - 3 / N^2]$$

$$A = [0.5(6-1)13.9519^4] / [13.9519(13.9519+1) - \{(2*13.9519-1)/6\} - 3/6^4] = 464.1268$$

$$K_{cc} = \sum X_i' / \sum Y_i' = 13.9519$$

Ahora bien, para que el valor de 13.9519 (K_{cc}) tenga validez, los valores de “ $1/K_i'$ ” debe ser estadísticamente independiente de los valores “ $\log m_i'$ ”. Para esto se realiza una regresión entre los valores de $1/K_i'$ (eje Y) y los valores de $\log m_i'$ como la eje X (Tablas 9 & 10).

Tabla 9. Los valores de “ $1/K_i'$ ” y “ $\log m_i'$ ”.

<i>I</i>	$1/K_i'$	$\log m_i'$
1	0.0820	1.87
2	0.0789	1.92
3	0.1238	1.62
4	0.0430	1.51
5	0.0547	1.68
6	0.0180	1.68

Tabla 10. Tabla de ANOVA para la regresión.

FV	Gl	SC	CM	F	F _{Tabla}
Reg.	1	0.02680	0.0268	16.75*	7.71
Error	4	0.00636	0.0016		
Total	5	0.03316			

Para tener independencia, la b (el coeficiente de la regresión = 0.0525) debe ser estadísticamente igual a cero. Esto se verifica mediante una prueba de t para la pendiente, en donde, la hipótesis nula y la hipótesis alterna son los siguientes:

$$H_0: b_{calculada} = 0$$

$$H_a: b_{calculada} \neq 0$$

$$t = (b_{calculada} - 0) / EE_b$$

$$EE_b = [CM_{Error} / SC_x]^{1/2}$$

Donde,

EE_b = Error estándar de la pendiente

CM_{Error} = Cuadro medio de error

SC_x = Suma cuadrado de los X 's

De los datos de la Tabla 10 se toma el valor de CM_{Error} lo cual es igual a 0.0016. Dado que el valor de $SC_x = 0.1231$, Se calcula el valor de Error estándar de la b mediante: $EE_b = [CM_{Error} / SC_x]^{1/2} = [0.0016 / 0.1231]^{1/2} = 0.1141$.

$$t = (b_{\text{calculada}} - 0) / EE_b = (0.0525 - 0) / 0.1141 = 0.4601$$

Ahora bien el valor calculado de t (0.4601) es menor del valor tabulado de t con $gl = 4$ a nivel de $\alpha = 0.05$ y bilateral (doble cola), lo cual es igual a 2.776. Por consiguiente, no se puede rechazar la hipótesis nula (H_0) y solo se puede apoyarla, es decir, los valores de “ $1/K_i$ ” (eje Y) son independiente de los valores de “ $\log m_i$ ” (eje X), y por tanto, se justifica el uso de K_{cc} .

2. Cálculo de K común ponderado (K_{cp})

Si por alguna razón, por ejemplo, la dependencia de los valores de $1/K_i$ en los valores de $\log m_i$, no se puede justificar el uso de un K común crudo, debemos proseguir con el cálculo y la justificación del K común ponderado.

$$K_{cp} = \sum W_i X_i^2 / \sum W_i X_i Y_i$$

Para esto, el investigador debe hacer los cálculos demostrados en la Tabla 8, y por medio de la siguiente ecuación calcular el valor de K_{cp} como sigue:

$$K_{cp} = 1474.385 / 113.9123 = 12.9432$$

Ahora se debe realizar una prueba de X^2 para verificar que el valor calculado de K_{cp} se ubica dentro del error aleatorio del muestreo. Se compara el valor calculado de X^2 ($X^2 = [\sum W_i Y_i^2 - (\sum W_i X_i Y_i)^2 / \sum W_i X_i^2]$), lo cual es igual a 1.3075, con el valor tabulado de X^2 (11.07) a nivel de $\alpha = 0.05$ y con $gl =$ número de zonas menos uno, es decir 6-1 igual a 5. Debido a que el valor calculado (1.3075) es menor que tabulado (11.07), se acepta que hay una justificación para el uso de K_{cp} .

Por tanto, el valor de $K_{cp} = 12.9432$, lo cual se puede utilizar para representar todas las 6 zonas. Ahora bien, existe una aplicación especial del uso de K_{cp} , en base a la ecuación.

$$\lambda = [M/2K_{cp}] * (Vc)$$

Donde, la M es la media de las 6 medias de las zonas bajo el estudio, $Vc =$ el valor de la tabla de X^2 a nivel de $\alpha = 0.05$ y con el $gl = 2K_{cp}$, y finalmente, la letra λ es un parámetro que tiene significancia particular, es decir, los valores de $\lambda < 2$ indican que la agregación se debe a los factores externos, mientras que los valores de $\lambda \geq 2$ significan que la distribución de tipo agregada se debe a los factores interiores y particularmente debido a la actividad de la población bajo del estudio.

En nuestro ejemplo, la $M = 55.1866$, el valor de $Vc = 38.885$, y por tanto, al substituir los datos en la ecuación $\lambda = [M/2K_{cp}] * (Vc)$, descubrimos que $\lambda =$

$55.1866/2(12.9432)](38.885) = 82.898$, el cual es mucho mayor que 2 y consecuentemente, la agregación se debe a los factores internos poblacionales.

Distribución espacial: Métodos alternos

Introducción

Existe un gran número de modelos sencillos propuestos por varios autores para determinar el tipo de dispersión espacial de las poblaciones; éstos métodos han sido utilizados en manejo integral de plagas (Binns & Nyrop, 1992, Badii et al., 1995, 1996b), en agricultura para las plagas insectiles (Badii & Moreno, 1992, Badii & Ortiz, 1992), ácaros plaga (Badii et al., 1994a, 1994b, Badii et al., 1996a, 1998), los depredadores (Badii & Flores, 1990, Badii & McMurtry, 1990), aquí se citan algunos de los modelos de uso más común en la literatura.

Índice de razón varianza-media

Esta prueba está basada en la igualdad entre la varianza y la media. La razón varianza-media o índice de dispersión ($I = v/m$) se aproxima a la unidad cuando la distribución es Poisson. Este índice frecuentemente se desvía de la unidad y la significancia de esta desviación se estima vía una prueba de $X^2 = (v/m)(n-1)$. Valores de I menor, igual o mayor de la unidad indican patrones de dispersión de tipo uniforme, Poisson o agregada, respectivamente. El parámetro de la dispersión o la K de binomial negativa se estima por medio de $K = m^2 / (v - m)$, y de aquí se despeja la v de forma siguiente: $v = m^2 / (K + m)$. Ahora bien la relación entre la K y el I es la siguiente: $I = v/m = [m^2 / (K + m)] / m = m/K + 1$. Puesto que $P = m/K$, por tanto, $I = P + 1$.

Índice de David & Moore

El índice de David & Moore (1954) se mide por: $I_{D\&M} = (v/m) - 1$. Valores de $I_{D\&M}$ menor, igual o mayor de cero indican patrones de dispersión de tipo uniforme, Poisson o agregada, respectivamente. En término de la relación con el parámetro de dispersión de binomial negativa o K , y en base a lo descrito arriba, $I_{D\&M} = p + 1 - 1 = p$, en otras palabras, ($I_{D\&M} = m/K$). Se utiliza el índice de David & Moore en dos casos siguientes:

1. Para comparar el grado de agregación entre dos poblaciones o para la misma población en dos localidades o tiempos distintos. Para esto hay que calcular las I_1 & I_2 según la ecuación $I_{D\&M} = (v/m) - 1$. Luego realizar una prueba de significancia para cada uno de estas dos índices para comprobar el grado de concordancia o diferencia entre las 2 índices.

$$\text{Prueba de significancia: } W = -\frac{1}{2} \ln (v_1/m_1 / v_2/m_2)$$

Si el valor calculado de W cae afuera del rango $(\pm 2.5/\sqrt{n-1})$, entonces las dos índices son estadísticamente diferentes a nivel de $\alpha = 0.05$ y viceversa.

2. Para determinar la naturaleza (tipo denso-independiente (DI), directamente denso-dependiente (DDD), inversamente denso-dependiente (IDD), etc.) de un factor de mortalidad, el índice de David & Moore funciona mejor que la K de binomial de negativa (BN) por el siguiente razonamiento.

Razonamiento

Un factor de mortalidad de tipo DI (denso-independiente) no altera el grado de agregación de una población cuya distribución es de tipo agregada y lo cual esta descrita via el modelo de binomial negativa (BN), esto significa que el valor del parámetro de dispersión de BN (K) debe quedar igual para la población inicial (antes del impacto de factor de mortalidad) y la población final, es decir, después del efecto de factor de mortalidad. Por el contrario, un factor de mortalidad de tipo DDD (directamente denso-dependiente) que apoya al balance natural, reduce el grado de contagio de una distribución agregada. Por tanto, vamos a comparar el $I_{D\&M}$ (y la ' K ' a partir de $I_{D\&M}$: $K=m/I_{D\&M}$) y la K de BN calculada por el tercer método del modelo de binomial negativa para este propósito según el ejemplo siguiente (Tabla 11).

Los datos (Tabla 11) indican que el factor de mortalidad destruye el 50% de la población ($\sum f_i X = 610$ versus $\sum f_f X = 305.02$) y de hecho reduce la media poblacional y también el valor de $I_{D\&M}$ a la mitad (3.05 vs 1.525, y 0.557 vs 0.278).

Tabla 11. Uso de K a partir de $I_{D\&M}$ y la K de BN.

X	f_i	f_f
0	40	63.59
1	30	44.06
2	10	40.78
3	20	31.88
4	30	15.16
5	40	4.06
6	30	0.47
	$\sum f_i = 200$	$\sum f_f = 200$
	$\sum f_i X = 610$	$\sum f_f X = 305.02$
Media (m)	$m_1 = 3.05$	$m_2 = 1.525$
Varianza (v)	$V_1 = 4.7475$	$V_2 = 1.9494$
$I_{D\&M}=(v/m) -1$	$(4.7575/3.05)-1=0.557$	$(1.9494/1.525) - 1 = 0.278$
$K=m/I_{D\&M}$	$3.05 / 0.557 = 5.4757$	$1.525 / 0.278 = 5.4856$
K: 3^{er} método	2.912	4.224
X = clase		
f_i = frecuencia inicial de la población		
f_f = frecuencia final después del impacto del factor de mortalidad		

También, se puede observar de la Tabla 11 que el valor de K calculada a partir del índice de David & Moore, prácticamente queda igual para las dos poblacionales inicial y final. Mientras que el valor de la K calculada a partir del tercer método de binomial negativa es mayor para la población final (4.224) comparado con la población inicial (2.912). En otras palabras, si utilizamos la K de BN, como el índice de dispersión espacial de estas dos poblaciones, llegaremos con la errónea conclusión que se trata de un factor de mortalidad de tipo directamente denso-dependiente (DDD), cuando en realidad los valores iguales de K calculadas a partir del índice de David & Moore demuestran que el factor de mortalidad remueve la población de forma aleatoria y por tanto, el factor de mortalidad es independiente de la densidad (DI) y no tiene ninguna injerencia sobre el equilibrio natural y esto tiene implicaciones fundamentales para el manejo de las poblaciones o recursos naturales y por ende, sobre la sustentabilidad. Se concluye que la K a partir del índice de David & Moore (contrario a la K de BN) tiene implicaciones y significancia ambiental y describe de forma concreta el uso racional o el abuso de los recursos naturales.

Modelo de Morisita

Debido a que el tamaño del cuadrante (UM) puede afectar la estimación del contagio, es deseable entonces tener una medida de dispersión independiente del tamaño de la unidad muestral (Morisita, 1959). Supóngase que la población consiste de manchones de individuos de diferentes densidades, y dentro de cada manchón los individuos están distribuidos de forma aleatoria; para situaciones como ésta se puede usar el modelo de Morisita (1959). $I_{\delta} = [Sn_i(n_i-1)/n(n-1)]N$, donde I_{δ} es el índice de Morisita, n_i es el número de individuos en "i" ésima unidad muestral, n es el número de individuos en todas unidades muestrales, y N es el número de unidades muestrales. Valores de I_{δ} menor, igual o mayor de 1 indican distribuciones de tipo uniforme, Poisson y agregada, respectivamente. La prueba de significancia para este índice es por medio de la siguiente ecuación: $F = [I_{\delta}(n-1) + (N-n)] / (N-1)$. El tamaño óptimo de la muestra se calcula a partir de: $n_{opt} = t^2 [(I_{\delta} + (1/m) - 1) / D^2]$, donde, t = valor de la tabal de t-student, m = la media, y D = grado de confiabilidad con valores de 0.25 para la aplicación y 0.10 para la investigación.

Ejemplo

Supongamos que los datos siguientes (Tabla 12) demuestran el número de los individuos en 17 unidades muestrales. La pregunta sería el determinar el patrón de distribución espacia y calcular el tamaño óptimo de la muestra.

Tabla 12. Número de los individuos (n_i) por cada unidad muestral (UM)

UM	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
n_i	2	7	9	4	1	0	9	10	13	6	1	7	8	9	1	2	5

Solución:

Cálculo de I_{δ} :

$$I_{\delta} = [Sn_i(n_i-1)/n(n-1)]N$$

$$= [\{2(1) + 7(6) + \dots + 5(4)\} / 94(93)] 17 = 1.2990$$

Pruebas de hipótesis para este ejemplo:

$$\mathbf{H_0: } I_{\delta} = 1$$

$$\mathbf{H_a: } I_{\delta} \neq 1$$

$$F_{\text{calculada}} = [I_{\delta}(n-1) + N-n]/(N-1) = [1.2990(94-1) + 17 - 94]/16 = 2.7379$$

Ahora bien, el valor tabulado de la F (F_{tabulada}) con gl 16 & ∞ es igual a 2.01. Valor de $F_{\text{calculada}}$ (2.7379) > valor de F_{tab} (2.01), por tanto, se rechaza la H_0 , es decir, la dispersión es agregada ya que el valor de $I_{\delta} > 1$.

El tamaño óptimo de la muestra sería:

$$n_{\text{opt}} = t^2 [(I_{\delta} + (1/m) - 1) / D^2]$$

$$\text{Para la aplicación: } 1.96^2 [1.299 + (1 / 5.5297)] / 0.25^2 = 3.325 \approx 4$$

$$\text{Para la investigación: } 1.96^2 [1.299 + (1 / 5.5297)] / 0.10^2 = 20.78 \approx 21$$

La ley de poder de Taylor

Taylor (1961) describió una relación potencial entre la media (m) y la varianza (v) muestral mediante la función de: $v = am^b$, donde " a " y " b " se estiman vía una regresión entre logaritmo de las varianzas y logaritmos de las medias de la forma siguiente: $\log v = \log a + b \log m$. La " a " es el antilogaritmo de la intersección con la ordenada y depende de la técnica de muestreo, tamaño de unidad muestral y tipo de hábitat; la " b " es la pendiente de la línea de regresión que determina el tipo de dispersión espacial, según Taylor, su modelo da consistentemente el mejor ajuste a los datos y además, ofrece los parámetros más confiables para describir el patrón de dispersión espacial de muchos organismos. El tamaño óptimo de muestra según este modelo es: $n = am^{(b-2)} / D^2$, donde, a' = antilogaritmo de $\log a$, D = grado de confiabilidad (0.25 para aplicación y 0.10 para la investigación) y todos los demás parámetros como antes descritos.

Cálculo del índice b de Taylor

En 1961, Taylor (Taylor, 1961) demostró que un número considerable de especies de diferentes tipos de organismos, tenían un elemento en común en su distribución espacial, que aunque no era constante en todas las densidades, sí se manifestaba en todas las

especies; una relación consistente entre la varianza (v) y la media de los datos de muestreo de forma siguiente $v = am^b$. Actualmente, se ha encontrado que muchas distribuciones estadísticas se combinan dentro de esta ley de “poder”, pero se ha mantenido el mismo índice, cuando se muestrean en diferentes ocasiones, la misma plaga y en el mismo cultivo. También el índice b , el coeficiente de regresión cuando la ley es expresada en logaritmos ($\log v = \log a + b \log m$), es dependiente de la interacción entre el comportamiento de la especie y representa las bases para un programa de muestreo. El método ha sido adoptado por la mayoría de los investigadores a nivel mundial, por ofrecer un buen equilibrio entre confiabilidad y accesibilidad práctica. En 1978, se aplicó para conocer la distribución de 102 especies animales y vegetales, y funcionó en forma convincente (Taylor et al., 1979). Se incluyeron en el análisis más de 200,000 unidades de muestra, y se comprobó una mejor descripción de la distribución espacial, que cuando se usó la media de hacinamiento (mean crowding) y la K de la binomial negativa. Ellos obtuvieron datos del encírtido *Oencyrtus kuvanae* parasitando los huevecillos de la palomilla *Lymantria dispar* en los bosques de Pensilvania y midieron la agregación en espacio y tiempo. Estos autores claramente, demostraron que se producen cambios secuenciales en la distribución de frecuencias, a medida que cambian las densidades dentro de la población; es decir, a medida que se manifiesta la ley de poder. La regresión varianza – media tuvo un bajo coeficiente (b) y una alta correlación (0.94); así mismo, la distribución estadística entre las muestras variaron desde el arreglo espacial de tipo Poisson a través de binomial negativa, con valores de K de 4.72 a 0.57 hasta la log-normal.

Esto significa que hubo un cambio continuo en la estructura básica de agregación, a medida que cambio la densidad. Este cambio gradual quedo representado por la función de poder $v = 1.08m^{1.27}$. En base a lo anterior, Taylor (1984) opina que las distribuciones estadísticas no tienen contenido biológico fundamental: ellas son únicamente descripciones de las fases por las que pasa, el incremento progresivo de asimetría y de la cola de la distribución, a medida que la densidad aumenta. Esto ya había sido citado por Anscombe en 1950, y después confirmado por Taylor con su Ley de Poder (Taylor, 1961). El método también fue aplicado por Badii et al., (1994b) usando un rango amplio de media de densidades.

Una vez que se obtiene el índice de agregación, el cual puede tomar valores según el tipo de distribución espacial ($b=1$, al azar; $b<1$, con tendencia a uniforme y $b>1$, en agregados) se puede calcular el tamaño óptimo de muestra con la formula:

$$n = a' m^{(b-2)} / D^2$$

Donde,

D = Error estándar como una fracción de la media.

a' = El intercepto de la regresión: $\log v = \log a + b \log m$ en forma antilogaritmo, es decir, a' = antilogaritmo de “ $\log a$.”

b = El parámetro que determina el tipo de la distribución espacial.

Modelo de Green

Green (1966) utilizó el índice de David & Moore (1954) para construir su modelo como sigue: $C_x = I_{D\&M} / (\sum X - 1)$, donde, $\sum X$ = número total de individuos en la muestra, lo cual es igual a Nm , donde, $N = \#$ de unidades muestrales, y la m = la media. Valores de C_x menor, igual o mayor de cero indican patrones de dispersión de tipo uniforme, Poisson o agregada, respectivamente.

Relación entre C_x y el parámetro K

Recordar que: $C_x = I_{D\&M} / (\sum X - 1) = m/K + 1 - 1/Nm - 1 = m/K/Nm - 1 = m/K / m(N - 1/m) = 1/K/N - 1/m$. De aquí podemos calcular la K derivado del modelo de Green y el tamaño óptimo de la muestra para el índice de Green: $C_x = 1/K/N - 1/m$, por tanto, $K = 1/N - 1/m(N - 1/m)$, de aquí, el tamaño óptimo de la muestra será: $N = (1/KC_x) + (1/m)$.

Elliott (1977) proporciona los valores para distribución de tipo Poisson y máxima uniformidad y máxima agregada (Tabla 13).

Tabla 13. Valores de varios índices de dispersión.

Índice	Max. uniforme	Poisson	Max. agregada
$I = V/m$	0	1	$\sum X$
$V(n-1)/m$	0	$n - 1$	$\sum X (n - 1)$
$I_{D\&M} = V/m - 1$	-1	0	$\sum X - 1$
Lexis ($\sqrt{V/m}$)	0	1	$\sqrt{\sum X}$
$(100\sqrt{(V - m)/ m})$ Charlier	Indefinido	0	$100n (1 - 1/\sum X)$
$C_x = I_{D\&M} / (\sum X - 1)$	$- 1/ (\sum X - 1)$	0	1
K (bin. Negativa)	$- m$	$\pm \infty$	$1/n [1 + 1/(\sum X - 1)]$
$1/K$	$-1/m$	0	$n - 1/m$
b (Taylor)	0	1	∞
I_{δ} (Morisita)	$1 - (n - 1)/(\sum X - 1)$	1	n
$\frac{\sum X^2}{(\sum X)^2}$ (Cole, 1946)	$1/n$	$1/n + (n-1)/n (1/\sum X)$	1
$1/45 \tan^{-1} (V/m) - 1$ (Lefkowitz, 1966)	-1	0	+1

Modelo de Lloyd

Hay que recordar que la media usual o de densidad (m) es en realidad el promedio de los individuos por cuadrante o unidad muestral, sin embargo, la media de hacinamiento (m^*) es el promedio de otros individuos por individuo por cuadrante. En otras palabras se tiene interés en saber qué pasa con el individuo bajo la observación o qué tan hacinado está este individuo y de aquí esta media obtiene un valor particular en término de la noción de competencia. Lloyd (1967) fue el primero que definió la m^* y lo estimó en base de $m^* = m + [(v/m) - 1]$. Este índice tiene la relación siguiente con el índice de David & Moore: $m^* = m + I_{D\&M}$. Lloyd denominó a la razón m^*/m el índice de patchiness, lo cual se relaciona con

la K de forma siguiente: $m^*/m = (m + I_{D\&M})/m$, o $m^*/m = 1 + I_{D\&M}/m$, este a su vez se convierte a: $1 + (m/K)/m = 1+1/K$, por tanto, el índice de patchiness es igual a $1 +$ inverso de K . Cabe mencionar que la media de hacinamiento depende en la media de densidad, sin embargo, el índice de patchiness es independiente de la densidad poblacional y solo obedece al patrón de distribución espacial.

Modelo de Iwao

Iwao (1968) en un estudio clásico encontró una relación lineal entre m y m^* de forma siguiente: $m^* = a + b m$, donde, a es el índice de contagio básico e indica el número de individuos que comparten la unidad de hábitat con un individuo a densidad infinitesimal y en realidad refleja la interacción innata entre los individuos. Valores de a igual a cero indican que la unidad bajo estudio es un individuo, mientras que valores mayores de cero significan que un grupo de individuos (por ejemplo, una colonia) forman la unidad del estudio; además, valores positivos y negativos de a son indicadores de la atracción y repelencia (competencia) entre los organismos, respectivamente. La b es el coeficiente de agregación de densidad e indica el tipo de dispersión espacial del organismo. Hay que mencionar que los valores numéricos (estadísticamente) de I_{δ} de Morisita (1959), b de Taylor (1961) y b de Iwao (1968) mayor, menor o igual a la unidad, indican dispersión espacial de tipo agregada, uniforme, o Poisson, respectivamente. Según Iwao, el tamaño óptimo de la muestra es lo siguiente: $n_{opt} = t^2[(\alpha + 1)/m + \beta - 1]/D^2$, donde, t es el valor de la tabla de t-student, α y β son los parámetros del modelo de Iwao, la m es la media de densidad, y la D es el grado de confiabilidad con valores de 0.25 para la aplicación y 0.10 para la investigación.

Una extensión de de la media de hacinamiento es la media de concentración o media de demanda, lo cual es igual a " m^*+1 ." Esta media involucra el individuo (el valor de 1 agregado a la media de hacinamiento) dentro del proceso de la utilización de los recursos. Es decir, la media de hacinamiento que era el promedio de *otros individuos* por cada individuo por cuadrante, se convierte en el promedio de *todos los individuos* (incluyendo el uno bajo del estudio) por cuadrante por individuo. En otras palabras, uno se compete, a parte con otros individuos, consigo mismo. Dicho de una manera simple, uno puede consumir lo que necesita para la mañana y por tanto, afecta su bienestar en futuro por su propia actividad de consumir el recurso escaso. Estos dos tipos de medias tienen aportaciones esenciales al concepto de la competencia. Hay que notar que la competencia tiene dos componentes. **1.** La *explotación*, lo cual significa reducir la probabilidad de disponibilidad de los recursos para los rivales, este componente refleja el efecto de estocasticidad en la competencia, de hecho es probable que los rivales no se conozcan uno al otro directamente, sin embargo, uno al utilizar el recurso finito (lo cual está en escases) reduce la probabilidad que este recurso sea disponible para su rival. **2.** La *interferencia*, lo cual se trata de combate físico entre los rivales en la competencia. Ahora bien, la media de demanda o concentración ($m^* + 1$) es el reflejo del componente de explotación en la competencia, y la media de hacinamiento (m^*) refleja la interferencia en la competencia.

Modelo de Pimbert & Jarry ($I_{(D)}$ o $I_{(SR)}$)

Según Pimbert & jarry (1988), supongamos que una vaine tiene n semillas numerados de pedúnculo hacia el extremo opuesta (1, 2, 3,..., n) y un insecto pone Y huevecillo en cada semilla (en total habrá p huevecillos por n semillas). Por tanto, $D = \sum Y_i^2$, nos ofrece una descripción buena de variabilidad de la dispersión de p huevecillos entre n semillas. Si las p huevecillos están distribuidas de forma al azar entre n semillas (hipótesis nula), la probabilidad de que un determinado huevecillo este depositado en la semilla “ i ” será $1/n$. Esta hipótesis se supone que. **1.** Cada semilla tiene la misma probabilidad de recibir cualquier huevecillo, **2.** Cada huevecillo tiene la misma probabilidad de ubicarse en cualquier semilla, y **3.** La presencia de un huevecillo en una semilla es independiente del otro huevecillo, es decir, la colocación independiente de huevecillos por semillas.

Bajo esta hipótesis, la media [$m_{(D)}$] y la varianza [$V_{(D)}$] de la D serán:

$$m_{(D)} = p/n (n + p-1)$$

$$V_{(D)} = 2p/n^2 (n-1) (p-1)$$

Por tanto, para N vainas, se puede usar el siguiente índice de dispersión:

$$I_{(D)} = (1/\sqrt{N}) \sum [(D_i - m_{(Di)}) / \sqrt{V_{(Di)}}]$$

Valores de:

- $I_{(D)} < 0$: significa una dispersión de tipo uniforme
- $I_{(D)} = 0$: significa una dispersión de tipo Poisson
- $I_{(D)} > 0$: significa una dispersión de tipo agregada

Recordar que: $n = \sum i$, y $p = \sum Y_i$

Ejemplo

Suponer que tenemos el siguiente arreglo de datos (Tabla 14).

Tabla 14. Datos para el cálculo de índice de Pimbert & Jarry.

n	1	2	3	4	5	6	7
Y	0	1	2	5	3	1	0

$$n = \sum i = 7$$

$$p = \sum Y_i = 0+1+...+0 = 12$$

$$D = \sum Y_i^2 = 0^2 + 1^2 + ... + 0^2 = 40$$

$$m_{(D)} = p/n (n + p-1) = 12/7 (7+12-1) = 30.85$$

$$V_{(D)} = 2p/n^2 (n-1) (p-1) = 2(12)/7^2 (7-1) (12-1) = 32.33$$

$$I_{(D)} = (1/\sqrt{N}) \sum [(D_i - m_{(Di)}) / \sqrt{V_{(Di)}}] = (1/\sqrt{1})(40-30.85)/\sqrt{32.33} = 1.6081$$

Por tanto, el patrón de dispersión espacial es agregado.

Este índice es bueno para estimar la desviación de la dispersión al azar, sin embargo, no puede distinguir entre los datos de la Tabla 14 y los de Tabla 15.

Tabla 15. Datos para el cálculo de índice de Pimbert & Jarry.

R	1	2	3	4	5	6	7
Y	2	5	3	1	1	0	0
Signo*	-	-	-	-	-	+	+
*: El signo “-” indica la semilla ocupada o infestada, y el signo “+” indica la semilla limpia o no infestada. R = Rango de las semillas.							

Al calcular el índice $I_{(D)}$ para los datos de la Tabla 15, obtenemos los resultados siguientes:

$$\begin{aligned}
 n &= 7 \\
 p &= 12 \\
 D &= 40 \\
 m_{(D)} &= 30.85 \\
 V_{(D)} &= 32.33 \\
 I_{(D)} &= 1.6081
 \end{aligned}$$

Estos resultados son exactamente iguales a los obtenidos de la Tabla anterior (14). Una exanimación de los datos de estos dos Tablas (14 y 15), revela que los datos son similares y la única diferencia se debe al arreglo o la secuencia de estos datos.

Por tanto, este índice ($I_{(D)}$) no es apropiado para distinguir la diferencias que ocurren en este o casos similares. Sin embargo, la diferencia entre estos arreglos de datos puede ser esencial para el bienestar de algunas especies. Imagínese que una especie de insecto prefiere depositar sus huevecillos en la parte basal de una vaina ya que a lo mejor esta posición le proporciona mayor protección a sus progenies y/o mayor calidad o cantidad de nutrientes para las larvas que emergen dentro de estas celdas o semillas. En otras palabras, puede haber preferencia diferencial para semillas distales, basales o centrales de una vaina con consecuentes ventajas para el desarrollo de la progenie. Por tanto, en situaciones de este tipo debemos contar con herramientas o índices apropiados que reflejen estas diferencias.

En este ejemplo, los datos de la Tabla 15 indican una preferencia de *no seleccionar* las semillas distales, es decir, las semillas número 6 y 7, algo que no se ve para los datos de la Tabla 14. Pimebert & Jarry (1988), precisamente desarrollaron un índice denominado el índice de serie de rango de Pimbert & Jarry ($I_{(SR)}$).

La ecuación de este índice es:

$$\begin{aligned}
 SR_i &= \sum R_j \\
 m_{(SR)} &= k(n+1) / 2 \\
 V_{(SR)} &= k(n-k)(n+1) / 2 \\
 I_{(SR)} &= (1/\sqrt{N}) \sum [(SR_i - m_{(SRi)}) / \sqrt{V_{(SRi)}}]
 \end{aligned}$$

Valores de:

- $I_{(SR)} < 0$: significa dispersión de tipo uniforme
- $I_{(SR)} = 0$: significa dispersión de tipo Poisson
- $I_{(SR)} > 0$: significa dispersión de tipo agregada

Donde,

“ R_j ” = El rango de semilla “j” marcado con el signo “+” y la “j” varia de 1 hasta “k”, es decir, “k” es la suma de las semillas vacías que llevan el signo “+”. Las semillas infectadas están marcadas con el signo negativo “-”. En el ejemplo de los datos de la Tablas 14 y 15, el cálculo de I_{SR} sería como sigue:

Datos de la Tabla 14:

$$\begin{aligned}
 k &= 2 \\
 SR &= \sum R_j = 1 + 7 = 8 \\
 m_{(SR)} &= k(n+1) / 2 = 2(7+1) / 2 = 8 \\
 V_{(SR)} &= k(n-k)(n+1) / 2 = 2(7-2)(7+1) / 2 = 40 \\
 I_{(SR)} &= (1/\sqrt{N}) \sum [(SR_i - m_{(SRi)}) / \sqrt{V_{(SRi)}}] = (1/\sqrt{1})(8 - 8) / \sqrt{40} = 0
 \end{aligned}$$

Datos de la Tabla 15:

$$\begin{aligned}
 k &= 2 \\
 SR &= \sum R_j = 6 + 7 = 13 \\
 m_{(SR)} &= k(n+1) / 2 = 2(7+1) / 2 = 8 \\
 V_{(SR)} &= k(n-k)(n+1) / 2 = 2(7-2)(7+1) / 2 = 40 \\
 I_{(SR)} &= (1/\sqrt{N}) \sum [(SR_i - m_{(SRi)}) / \sqrt{V_{(SRi)}}] = (1/\sqrt{1})(13 - 8) / \sqrt{40} = 0.79
 \end{aligned}$$

Por tanto, la dispersión de los datos de la Tabla 14 ($I_{(SR)} = 0$) es Poisson, mientras que para los de la Tabla 15 ($I_{(SR)} = 0.79$), es agregada.

Ejemplos para estimar la distribución espacial

Utilizaremos los datos de la Tabla 1 de esta sección, para ilustrar la determinación del patrón de distribución espacial (Tabla 16) para los índices siguientes:

- I:** Relación varianza-media = V/m
- $I_{D\&M}$:** Índice de David & Moore = $(V/m - 1)$
- Lex:** Índice de Lexis = $(\sqrt{V/m})$
- Ch:** Índice de Charlier = $(100\sqrt{(V - m) / m})$
- Green:** Índice de green = $(C_x = I_{D\&M} / (\sum X - 1))$
- 1/K:** Índice o parámetros de binomial negativa
- Cole:** Índice de Cole = $(\sum X^2 / (\sum X)^2)$
- Lef:** Índice de Lefkovitch = $(1/45 \tan^{-1} (V/m) - 1)$
- I_s :** Índice de Morisita = $[S n_i(n_i-1) / n(n-1)]N$

Tabla 1. Número de personas por oficina por cada empresa.

Empresas	Oficinas									
	1	2	3	4	5	6	7	8	9	10
1	3	1	1	3	1	3	2	4	1	2
2	4	3	1	2	2	0	3	2	2	2
3	2	3	1	2	2	5	1	3	2	4
4	3	3	1	1	2	2	4	4	2	0
5	2	1	1	1	2	1	4	4	0	3
6	2	2	3	1	1	2	2	2	1	0
7	2	2	1	1	4	1	1	0	1	1
8	1	0	0	4	1	1	3	2	1	3
9	3	4	3	0	2	1	4	4	3	1
10	1	5	3	3	2	0	0	0	1	2

Se calcula cada uno de estos índices para los datos (Tabla 1) de cada uno de las 10 oficinas, y para cada empresa singular (Tabla 16).

Tabla 16. Uso de índices para determinar la distribución espacial.

I	Oficinas										Empres
	1	2	3	4	5	6	7	8	9	10	
<i>V/m</i>	.983	.621	.503	2.193	.889	.577	.681	1.00	.293	.583	.859 U*
<i>I_{D&M}</i>	-.017	-.379	-.497	1.193	-.111	-.423	-.319	0.00	-.097	-.417	-.141 U
<i>Lex.</i>	1.043	.788	.709	1.481	.943	.759	.825	1.00	.293	.765	.927 U
<i>Ch.</i>	18.85	58.14	Ind.*	34.19	Ind.*	Ind.*	Ind.*	Ind.*	Ind.*	Ind.*	.767 U
<i>C_x</i>	.004	-.036	-.025	-.066	-.006	-.021	.014	0	-.071	-.197	.001 P
<i>1/K</i>	.918	.933	.481	.628	-.056	-.002	.142	.227	.643	.643	3.035 A
<i>Cole</i>	.139	.183	.102	.158	.140	.124	.127	.143	.123	.124	.014 U
<i>Lef.</i>	-.43	.044	-.413	.491	-.087	-.341	-.247	0.00	-.641	.332	-.605 U
<i>I_δ</i>	1.003	1.333	.762	1.111	.947	.809	1.00	.364	.823	.826	.987 U

: I = índice, Ind. = indefinido, U = uniforme, P = Poisson, A = Agregada, Empres = empresa

En término general y en base a los límites de valores para determinar el tipo de distribución espacial, se concluye que la mayoría de los casos, casi todos los índices (con excepción de *C_x* y *1/K*) indican el patrón de distribución espacial es de tipo uniforme. En caso del índice de Charlier, la palabra Ind* significa indefinido o imaginario, es decir, este índice no produce un valor preciso para el patrón de distribución espacial de tipo uniforme.

Además se puede determinar el tipo de distribución espacial para todos los datos de la Tabla 1 por medio del modelo de Taylor (1961).

Acordar la ecuación del modelo de Taylor:

$$V = a m^b$$

$$\log V = \log a + b \log m$$

De los datos de la Tabla 1 se arroja los parámetros (a) y (b):

$$a = 0.3991$$

$$b = 1.9456$$

Por tanto:

$$V = 0.3991 m^{1.9456}$$

Tabla 17. Tamaños óptimos de la muestra para diferentes medias.

Empresa	m	n_{opt} ($D = 0.25$)	n_{opt} ($D = 0.10$)
1	1.1	7	40
2	1.7	7	39
3	1.9	7	39
4	2.0	7	39
5	2.1	7	39
6	2.1	7	39
7	2.1	7	39
8	2.2	7	39
9	2.3	6	38
10	2.5	6	38

Y en base a la ecuación para el tamaño óptimo de la muestra para el modelo de Taylor: $n_{opt} = a m^{(b-2)} / D^2$, se calculan los tamaños óptimos de la muestra para diferentes medias (Tabla 17) para el caso de aplicación ($D = 0.25$) y la investigación ($D = 0.10$).

De los datos de la tabla arriba, deducimos que: **1.** los tamaño óptimos de la muestra se disminuyen a media que se incrementa el valor de la media muestra, es decir una relación inversa entre la m y el n_{opt} , **2.** los tamaños óptimos de la muestra para el caso de la investigación son casi 6 veces más grandes comparados con el caso de la aplicación.

Selección de un índice perfecto

Según Green (1966), Lefkovitch (1966) y Taylor (1984) un índice perfecto para la determinación de la distribución espacial es aquel índice que tiene los siguientes rasgos. Badii et al. (1994_a, 1996a) han reportado la forma correcta de seleccionar un índice bueno de distribución espacial.

1. Provee valores reales y continuos a lo largo de diferentes tipos de distribución espacial.
2. Debe ser independiente del tamaño poblacional, la media poblacional y el tamaño de la muestra.

3. Que sea una función de la varianza muestral y el parámetro k de binomial negativa.
4. Que tenga una prueba de significancia.
5. Que tenga aplicaciones prácticas.

Distribución normal y transformación

Si existe igual probabilidad de que un evento ocurre de una u otra manera en una distribución binomial ($p = q = 0.5$) y la el parámetro de K se aproxima a infinito, entonces, las series de probabilidad dadas por $(p+q)^K$ se aproximan a una curva simétrica y suave en forma de campana, lo cual es esencialmente la curva normal. La distribución normal esta asociada con las variables continuas; las mediciones en lugar de los conteos. Esta distribución raramente es apropiada para los conteos, sin embargo, es relevante, debido a que un gran número de los métodos (ANOVA, t, regresión, etc.) están asociados a esta curva. El uso de estos métodos estadísticos requiere las condiciones siguientes:

1. Los datos deben tener una distribución normal.
2. La varianza de la muestra debe ser independiente de la media.
3. Los componentes de la varianza deben ser aditivos.

La distribución de binomial positiva o uniforme es aproximadamente normal si el número de las unidades muestrales es grande (>30) y la varianza de la muestra es no menor de 3 ($V = pqK$). Por tanto, se puede usar la aproximación a normalidad cuando $p = 0.4$ a 0.6 para K entre 10 y 30, o cuando $p = 0.10$ a 0.9 para $K > 30$; y no se puede usarla cuando $K < 10$.

En la serie de Poisson, la distribución de frecuencias es asimétrica para valores pequeñas de K , pero se aproxima a normalidad a medida que el valor de λ incrementa en tamaño, y es normal cuando $\lambda > 10$.

La distribución de binomial negativa es asimétrica para un rango amplio de la media (m) cuando la K es pequeña (por ejemplo, $K = 2.5$), pero se aproxima a normalidad cuando el valor de K se incrementa y la media (m) es razonablemente grande (por ejemplo, $m = 10$). Cuando la K se aproxima a infinito, la distribución es idéntica a la de Poisson.

Por tanto, se reúna la primera condición de normalidad para las 3 distribuciones (uniforme, Poisson y binomial negativa).

Sin embargo, debido a que la varianza y la media tienden a incrementar conjuntamente en los 3 distribuciones mencionadas, la segunda condición (la independencia de varianza de la media) nunca se reúne. Consecuentemente, algunos métodos como la de t-student, ANOVA, no se pueden aplicar sin riesgo de errores considerables. Se puede remediar esta dificultad mediante el reemplazo de cada conteo por una función adecuada, por ejemplo logaritmo de los conteos. En otras palabras, los conteos se transforman, y la transformación correcta debe normalizar la distribución de frecuencias de los conteos, eliminado la dependencia de las varianzas sobre las medias, y asegurar que los componentes de la varianza sean aditivos.

La selección de la transformación correcta depende de la distribución original de frecuencias de los conteos (Tabla 16). Cabe destacar que se puede realizar las pruebas de t-student, ANOVA, correlación y regresión sobre los datos transformados. Cuando se termina el análisis, las estimaciones, por ejemplo, la media se debe transformarse a los datos originales, es decir, para la transformación de tipo de raíz cuadrada, la media transformada (la media derivada) al final del análisis debe elevarse a cuadrado, etc.

Debido a que el valor de la media derivada es más pequeño que la media original de los conteos (antes de transformación), no se puede comparar la media derivada con la media original. Por tanto, se debe hacer ajustes pequeños a las medias derivadas:

1. Transformación por raíz cuadrada: Se eleva la media derivada al cuadrado, se sustrae 0.5 del valor obtenido (si es el caso), y se agrega a la media derivada la varianza de la media derivada.
2. Transformación por logaritmo: Antes de sacar el antilogaritmo de la media derivada, se multiplica la varianza de los datos transformados por 1.15 y se agrega el valor resultante a la media derivada. Luego se obtiene el antilogaritmo del valor resultante.

El valor final, es decir, la media derivada más el ajuste, usualmente concuerda con el valor de la media obtenida por sacar la media de forma directa.

Tabla 16. Transformaciones adecuadas para normalizar los conteos.

Distribución original	Distribución desconocida	Condiciones especiales	Transformar la X por:
Poisson	$V = m$	Ningún $X < 10$	\sqrt{X}
	$V = m$	Algunos $X < 10$	$\sqrt{(X+0.5)}$
Bin. positive	$V < m$	$K > 5$	$\text{Sinh}^{-1}\sqrt{(X+0.375)/(K-2(0.375))}$
Binomial Negativa	$V > m$	K entre 2 y 5	$\log(X + K/2)$
	$V > m$	Ningún $X = 0$	$\log X$
	$V > m$	Algunos $X = 0$	$\log(X+1)$

La familia de distribuciones binomiales

La relación entre varios miembros de la familia de distribuciones de tipo binomial se resume de la siguiente forma (Tabla 17).

Tabla 17. Relación entre los miembros de familia binomial.

Familia binomial		
Binomial positiva ($V < m$) $(q + p)^K$	Binomial negative ($V > m$) $(q - p)^{-K}$	
$K \rightarrow + \infty$	$K \rightarrow - \infty$	$K \rightarrow 0$
Normal "V" independiente de "m" $q = p = 0.5$	Poisson $V = m$ $P \rightarrow 0$ $q \rightarrow 1$	Serie logarítmica

Conclusiones

Realizar el muestreo de forma objetiva y con rigor científico requiere el conocimiento del patrón de distribución de frecuencias de las mediciones o conteos. Existen tres tipos generales de distribución de frecuencias: **1.** El de tipo al azar que se divide en: **a.** la distribución normal en donde la el parámetro de K tiende hacia el infinito, las probabilidades de éxito (p) y el de fracaso (q) son iguales a 0.5 cada uno, y la varianza de la muestra es independiente de la media muestral, y **b.** distribución Poisson en donde la el parámetro de K tiende hacia el infinito, la probabilidad de éxito (p) se aproxima a cero y el de fracaso (q) tiende a uno, y la varianza y la media muestral son iguales. **2.** Distribución de tipo uniforme con $(q + p)^K$ y la varianza muestral es menor de la media muestral. **3.** Distribución de tipo agregada con $(q - p)^K$, la varianza muestral es mayor de la media muestral y el valor de K es menor que 2 de tal forma que valores fraccionales de K (aproximarse a cero) constituyen la extrema agregación representada por las distribuciones de tipo de serie logarítmica y logaritmo normal.

Referencias

- Anscombe, F.J. 1950. Samplin theory of negative binomial logarithmic series distributions. *Biometrika*, 37: 358-381.
- Badii, M.H. & A. E. Flores. 1990. Ecological studies of mites on citrus in Nuevo Leon, Mexico: Preliminary surveys for phytoseiids. *Internat. J.Acarol.* 16 (4): 235-239.
- Badii, M.H. & J.A. McMurtry. 1990. Field experiments on predation, dispersion, regulation and population changes. *Publ. Biol.* 4(1-2): 43-48.
- Badii, M.H. & M.D. Moreno. 1992. Patrón de dispersión espacial y fluctuación poblacional de tres especies de barrenadores del fruto del tomate. *Publ. Biol.* 6(2): 184-188.
- Badii, M.H. & M.C. Ortíz. 1992. Análisis de dinámica poblacional y dispersión espacio-temporal del picudo del chile sobre chile Jalapeño. *Publ. Biol.* 6(1): 61-64.
- Badii, M.H., A.E. Flores, S. Flores & S. Varela. 1994a. Statistical description of population distribution and fluctuation of citrus rust mite (Acari: Eriophyidae) on orange fruit in Nuevo Leon, Mexico. *Biotam*, 6(1): 1-8.
- Badii, M.H., A.E. Flores, S. Flores & S. Varela. 1994b. Comparative estimation of distribution statistics of citrus rust mite (Acari: Eriophyidae) on leaves of three different orange orchards in Nuevo León, Mexico. *Biotam*, 6(1): 9-16.
- Badii, M.H., A.E. Flores, R. Torres & H. Quiroz. 1995. Muestreo y evaluación económica de las plagas. Pp. 1-13. En: H. Fuente (ed.). *Curso Internacional sobre Manejo de Huertas de Cítricos*. SAGAR, INIFAP, CIRNE., General Allende, N. L., México.
- Badii, M.H., A. E. Flores, S. Varela, S. Flores & R. Foroughbakhch. 1996a. Dispersion indices of citrus rust mite (Acari: Eriophyidae) on orange in Tamaulipas, Mexico. Pp. 17-20. En: R. Mitchel, D. Horn, G. R. Needham y W. C. Welbourn (eds.). *Acarology IX: Volume 1, Proceedings. Section I: Behavior and Physiological Ecology*. Ohio Biological Survey, Columbus, Ohio.
- Badii, M.H., A.E. Flores, R. Foroughbakhch, H. Quiróz & R. Torres. 1996b. Ecología de manejo integrado de plagas (mip) con observaciones sobre control microbiano de insectos. Pp. 21-49. En: L.J. Galan Wong, C. Rodreiguez-Padilla y H. Luna-Olvera (eds.). *Avances Recientes en la Biotecnología en Bacillus Thuringiensis*. Ciencia, Universitaria no. 2. UANL.

- Badii, M.H., A.E. Flores, R. Foroughbakhch & H. Quiroz. 2000. Fundamentos de muestreo. Pp. 129-144. In: M.H. Badii, A.E. Flores & L.J. Galán (eds.). Fundamentos y perspectiva de Control Biológico, UANL, Monterrey.
- Badii, M.H., F. López Pérez, A.R. Pazhakh & K. Sáenz López. 2004. Muestreo como un requisito fundamenta len las ciencias experimentales. *InnOvaciOnes de NegOciOs*, 1(1) 33-53.
- Badii, M.H., J. Castillo & A. Wong. 2006. Diseños de distribución libre. *InnOvaciOnes de NegOciOs*, 3(1): 141-174.
- Badii, M.H., J. Castillo, A. Wong & J. Landeros. 2007a. Precisión de los índices estadísticos: técnicas de jackknife & bootstrap. *InnOvaciOnes de NegOciOs*. 4(1): 63-78.
- Badii, M.H., J. Castillo, J. Landeros & K. Cortez. 2007b. Papel de la estadística en la investigación científica. *InnOvaciOnes de NegOciOs*. 4(1): 107-145.
- Badii, M.H., J. Castillos, R. Foroughbakhch & K. Cortez. 2007c. Probability and scientific research. *Daena*, 2(2): 358-369.
- Badii, M.H., J. Castillo & A. Guillen. 2008. Tamaño óptimo de la muestra. *InnOvaciOnes de NegOciOs*, 5(1): 53-65.
- Badii, M.H. & J. Castillo. 2008. Una tabla estadística para los K-valores. *CULCYT*, 5(29): 24-29.
- Badii, M.H. & J. Castillo. 2009. Distribuciones probabilísticas de uso común. *Daena*, 4(1): 149-178.
- Badii, M.H. & J. Castillo. 2010. Muestreo Estadístico: Conceptos y Aplicaciones. UANL, Monterrey. 250 pp.
- Badii, M.H. & A. Guillen. 2010. Decisiones estadísticas: bases teóricas: *Daena international J. Good Conscience*. 5(1): 185-207.
- Badii, M.H., L.A. Araiza & A. Guillen. 2010a. Esenciales de la estadística: Un acercamiento descriptivo. *Daena intenernational J. Good Conscience*. 5(1): -208-236.
- Badii, M.H., A. Guillen & L.A. Araiza. 2010b. Estimaciones estadísticas: Un acercamiento analítico. *Daena international J. Good Conscience*. 5(1): 237-255.
- Badii, M.H., A. Guillen & J.L. Abreu. 2010c. Representatividad estadística en las ciencias sociales. *Daena International J. of Good Conscience*. 5(2): 170-218.
- Binns, M. R. and J. P. Nyrop. 1992. Sampling insect populations for the propose of IPM decision making. *Ann. Rev. Entomol.* 37: 427-453.
- Bliss, C.I. & R.A. Fisher. 1953. Fitting the negative binomial distribution to biological data and note on the efficient fitting of the negative binomial. *Biometrics*, 9: 176-200.
- Bliss, C.I. & A. R.G. Owen. 1958. Negative binomial distribution with a common K. *Biometrika*, 45: 37-58.
- Cole, L.C. 1946. A theory for analyzing contagiously distributed populations. *Ecology*, 27: 329-341.
- David, F.N. & P.G. Moore. 1954. Note on contagiousness distributions in plant populations. *Ann. Bot.* 18: 47-53.
- Elliott, J.M. 1977. Statistical Analysis of Samples of Benthic Invertebrates. Freshwater Biological Association. Cambria. Luisiana. USA.
- Green, R.H. 1966. Measurement of non-randomness in spatial distributions. *Res. Popul. Ecol.* 8: 1-7.
- Iwao, S. 1968. A new regression method for analyzing the aggregation pattern of animal populations. *Res. Popul. Ecol.* 10: 1-20.
- Lefkovitch, L.P. 1966. An index of spatial distribution. *Res. Popul. Ecol.* 8: 89-92.
- Lloyd, M. 1967. Mean crowding. *J. Anim. Ecol.* 36: 1-30.
- Morisita, M. 1959. Measuring the dispersion of individuals and analysis of the distributional patterns. *Mem. Fac. Sci. Kyushu Univ. Ser. E.* 2: 215-235.
- Pimpbert, M.P. & M. Jarry. 1988. A non-parametric description of the oviposition pattern of *Zabrotus subfasciatus* inside pods of a wild, *Phaseolus lunatus*, and a cultivated host plant, *Phaseolus vulgaris*. *Insect Sci. Appl.* 9(1): 113-116.
- Taylor, L.R. 1961. Aggregation, variance and the mean. *Nature*, 189:732-735.

- Taylor, L.R. 1984. Assessing and interpreting the spatial distributions of insect populations. *Ann. Rev. Entomol.* 29: 321-257.
- Taylor, L.R., I.P. Woiwood & N. Perry. 1979. The density dependence of spatial behaviour and the rarity of randomness. *J. Anim. Ecol.* 47: 383-406.
-

***Acerca de los Autores**

El Dr. Mohammad Badii es Profesor e Investigador de la Universidad Autónoma de Nuevo León. San Nicolás, N.L., México, 66450. mhbadiiz@gmail.com

La Dra. Amalia Guillén es Profesora-Investigadora, Egresada de FACPYA-UANL. Monterrey, NL.

El Dr. E. Cerna es Profesor e Investigador en UAAAN, Saltillo, Coah.

El Dr. J. Landeros es Profesor e Investigador en UAAAN, Saltillo, Coah.