

Distribución Espacial

Eduardo Chacón-Madrigal

2021-09-16

Distribución espacial

En la naturaleza existe variación en cómo se distribuyen los organismos en el espacio o en cómo ocurren ciertos eventos en el tiempo. Describir esos patrones y explicarlos es una tarea de los ecólogos. Estos distintos patrones de distribución tienen implicaciones en la vida de los organismos o son estrategias ante la variación del ambiente. Aunque la variación en los patrones de distribución es continua, se pueden describir tres patrones de distribución: uniforme, aleatorio o agregado (Fig. 1). Por ejemplo, los elefantes se encuentran distribuidos de manera agregada porque viven en grupos y mantienen relaciones sociales muy estables y algunos árboles se distribuyen de manera uniforme porque compiten entre ellos.

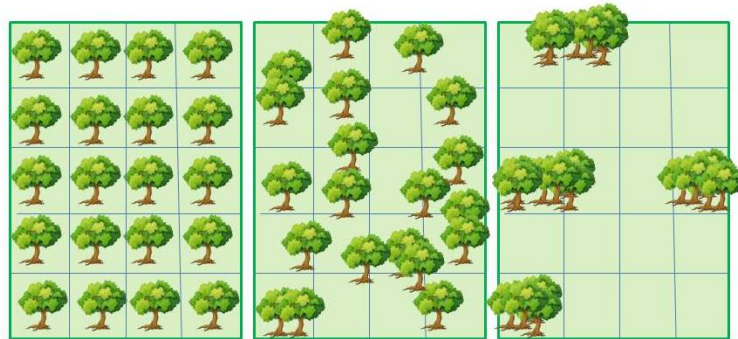


Fig. 1 Tipos de Distribución espacial Uniforme, Aleatorio y Agregado.

Los patrones de distribución espacial nos dan mucha información sobre la historia de vida de los organismos y son necesarios de conocer si queremos manejar las poblaciones. Para describir esos patrones existen diferentes métodos. En una situación ideal uno podría tener mapeados a todos los individuos de la población en estudio y aplicar métodos de análisis más sofisticados y pueden dar otra información como la densidad de población. Sin embargo, eso es difícil de obtener y normalmente se trabaja con datos obtenidos en cuadrantes u observaciones en transectos.

Métodos para mapas espaciales.

El vecino más cercano

En algunas ocasiones se podría tener un mapa exacto de la ubicación de los individuos de la población en estudio. Supongamos una situación como en el segundo cuadro de la figura 1; en ese caso se podría tener la

distancia entre individuos. Se puede estimar la distancia promedio al individuo más cercano con la siguiente fórmula:

$$\bar{r}_A = \frac{\sum r_i}{n}$$

En donde

$$r_A$$

, es el promedio de distancia al vecino más cercano.

$$r_i$$

, es la distancia al vecino más cercano de cada individuo y

$$n$$

, es el número de individuos. Si se tiene el área de estudio entonces se puede calcular la densidad

$$\rho$$

, con la fórmula:

$$\rho = \frac{n}{area}$$

En donde n es el número de individuos y área es el área de estudio. Con la densidad se puede calcular la distancia promedio al vecino más cercano esperada. La distancia esperada es suponiendo que los individuos están dispersos de manera aleatoria.

$$\bar{r}_E = \frac{1}{2\sqrt{\rho}}$$

Con estos valores se puede calcular un índice de agregación R, que es una proporción entre la distancia promedio observada y la distancia promedio esperada:

$$R = \frac{\bar{r}_A}{\bar{r}_E}$$

Si el patrón de distribución espacial es aleatorio entonces R=1. Si es agregado entonces R tiende a 0 y si el patrón es uniforme R aproxima a un límite de 2.15. Para poder probar estadísticamente si R es diferente de 1 se utiliza una prueba Z.

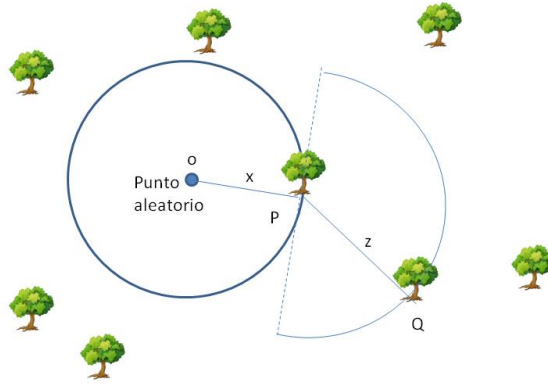
$$Z = \frac{\bar{r}_A - \bar{r}_E}{s_r}$$

Sr es un error estándar de la distancia al vecino más cercano esperada.

$$s_r = \frac{0.26136}{\sqrt{n\rho}}$$

Método de distancia

En algunos casos no es posible ubicar en el espacio a toda la población sino únicamente individuos dispersos en una gran área. Para eso se han propuesto varios procedimientos de campo y varios tipos de análisis. Uno de los métodos propuestos se conoce como muestreo en T (Fig. 2). En el muestreo en T se establece un punto al azar y a partir de ese punto se mide la distancia al individuo más cercano. Desde el individuo más cercano se busca el próximo más cercano en la dirección opuesta al punto aleatorio, que nunca debe estar en un ángulo menor a 90 grados con respecto al punto aleatorio y el primer individuo más cercano (Fig. 2).



$$h_t = \frac{2n[2\sum(x_i^2) + \sum(z_i^2)]}{[(\sqrt{2}\sum x_i) + \sum z_i]^2}$$

Ht es el estadístico para aleatoridad de datos de muestreo en T.

n= tamaño de muestra (número de puntos).

xi=distancias a los puntos.

zi= distancias entre los organismos.

El valor Ht=1.27 indica una distribución aleatoria, valores menores indican una distribución uniforme y valores mayores indican una distribución agregada. Para definir si el valor Ht es mayor o menor, el mismo ht funciona como un estadístico y se utiliza una tabla ajustada a este valor.

Cuadro 1. Valores críticos del Ht que prueba como hipótesis nula si la distribución es aleatoria; se presenta los valores tabulares para el alfa de 0.05. Por ejemplo, si ht=1.6 en una muestra de 5 puntos entonces la distribución es agregada con un p>0.05.

	Uniforme	Agregada
n/alfa	0.05	0.05
5	1.09	1.4593
6	1.11	1.4472
7	1.12	1.4368
8	1.13	1.428
9	1.13	1.4203
10	1.14	1.4136

	Uniforme	Agregada
11	1.15	1.4078
12	1.15	1.4025
13	1.16	1.3978
14	1.16	1.3936
15	1.16	1.3898
17	1.71	1.3815
20	1.77	1.3748
25	1.19	1.3644
30	0.19	1.3565
35	1.20	1.3504
40	1.20	1.3455
45	1.21	1.3414
50	1.21	1.3379
75	1.22	1.326
100	1.23	1.3189
150	1.24	1.3105
200	1.24	1.3055
300	1.25	1.2995
400	1.25	1.296
500	1.25	1.2936

Métodos de cuadrantes

Para describir los patrones de distribución también se han utilizado métodos de cuadrantes o unidades de muestreo. En estos métodos se cuentan los organismos dentro de cada unidad de muestreo. Este tipo de métodos también sirve para ver agrupaciones de eventos en unidades de tiempo. Por ejemplo, una unidad de muestreo puede ser un área de tamaño determinado, una estructura específica (ej. una planta, una flor) o intervalos de tiempo de igual tamaño. Cuando se trata de individuos dispersos en el espacio generalmente se hacen parcelas en donde se cuentan los individuos de una o varias especies. Cada parcela es una unidad de muestreo. Si la distribución del organismo en estudio es aleatoria, o al azar, entonces la probabilidad de encontrar un individuo es la misma en cualquier punto, independiente de la presencia de otro. Si la distribución es agrupada entonces la probabilidad de encontrar un individuo aumenta si hay otro. Finalmente, si la distribución es uniforme entonces la probabilidad de encontrar un individuo disminuye cuando ya se ha encontrado otro. El patrón más común de distribución es el agregado.

Para estudiar los patrones se utilizan los índices de dispersión. El índice de dispersión más sencillo de entender es la relación entre la variancia y el promedio.

$$CD = \frac{s^2}{\bar{x}}$$

Veamos un ejemplo: supongamos que en cada una de las parcelas de la figura 1 hay 20 árboles y se hicieron 20 sub parcelas. El resultado final es que en cada una de las parcelas hay en promedio 1 individuo por sub parcela. Sin embargo, entre parcelas lo que cambia es la variancia, por ejemplo, en el primer caso, la variancia tiende a ser igual que el promedio, o sea en este caso el índice de dispersión es igual a 1. En el segundo caso la variancia tiende a 0, o sea en este caso el índice de dispersión es menor que 1. En el último caso, la variancia es muy alta porque hay sub parcelas sin árboles y otras con muchos, por lo tanto el índice de dispersión es mayor que 1.

Usemos un ejemplo para entender el Índice de Dispersión

Creamos un conjunto de datos

El conjunto de datos es la cantidad de individuos de la palma *Chamaedorea* en parcelas de 2 x 2 m en la Reserva Leonelo Oviedo.

```
#Creamos un conjunto de datos (estos datos son parte de una práctica de Ecología con palmas del género Coccothrinax)
individuos<-c(9,10,7,14,4,7,3,20,6,3,4,9,10,4,10,1,6,7,5,7,4,4,1,2,7,8,6,5,9,11,9,9,10,13,16,4,14,12,7,9)
parcelas<-seq(1, length(individuos), 1)
datos<-data.frame(parcelas, individuos)
```

Exploremos los datos

En el conjunto de datos se hicieron 80 parcelas y dentro de cada parcela se contaron los individuos.

```
head(datos, 10) # Visualizar los primeros 10 datos
```

##	parcelas	individuos
## 1	1	9
## 2	2	10
## 3	3	7
## 4	4	14
## 5	5	4
## 6	6	7
## 7	7	3
## 8	8	20
## 9	9	6
## 10	10	3

#Calculando el Índice de dispersión Ahora teniendo del número de individuos por parcela calculamos el índice de dispersión:

$$CD = \frac{s^2}{\bar{x}}$$

Creamos la función para calcular el índice de dispersión usando la función de la variancia **var()** y la función del promedio **mean()**. Podemos especificar que no tome en cuenta las casillas en las que se perdieron datos por alguna razón, con el parámetro **na.rm=TRUE**

```
CoefDisp<-function(x){var(x, na.rm=TRUE)/mean(x, na.rm=TRUE)}
CD<-CoefDisp(datos$individuos)
paste("Mi índice de dispersión es:", CD, sep="")
```

```
## [1] "Mi índice de dispersión es:2.28743649358478"
```

¿Es mi Índice de dispersión estadísticamente diferente de uno?

Si tenemos pocas parcelas el índice puede variar mucho y no sabríamos si por azar el valor obtenido es alto, bajo o igual a uno, es por eso que se requiere una prueba. Para saber eso usamos una prueba de

chi cuadrado multiplicando el índice de dispersión por el número de grados de libertad que corresponde al número de parcelas menos uno.

$$\chi^2 = CD(n - 1)$$

```
# Crea la función para calcular el chi cuadrado del índice de dispersión
ChiCD<-function(x){CoefDisp(x)*(length(x)-1)}

# Luego calculamos el chi
mi.chiCD<-ChiCD(datos$individuos)

# Calculamos los grados de libertad
mis.gl<-length(datos$individuos)-1

# Calculamos la probabilidad para el chi correspondiente
mi.p<-dchisq(x=mi.chiCD, df=mis.gl)

#Ponemos todo junto
paste("Mi Coeficiente de Dispersión es:", round(CD,3),
      "; Chi2 =", round(mi.chiCD,3),
      "; g.l.=", mis.gl, "; p=",
      round(mi.p,5), sep="")
```

```
## [1] "Mi Coeficiente de Dispersión es:2.287; Chi2 =180.707; g.l.=79; p=0"
```

```
# NOTAS: la función round(), sirve para redondear
# La función paste() sirve para pegar objetos con texto
```

Interpretación

La prueba indica que el coeficiente de dispersión es estadísticamente mayor a 1 y por lo tanto la distribución es agregada. Hay que tener cuidado que las pruebas de chi cuadrado que generalmente se aplican son de una cola. En este caso interesa si el valor es menor o mayor a 1, por lo tanto es de dos colas. Como es una prueba de dos colas, para aceptar la hipótesis nula con una probabilidad alfa de 0.05, el valor de chi cuadrado obtenido debe estar entre los valores de $p = 0.975$ y 0.025 .

Otras formas de comparar las distribuciones

El índice de dispersión es fácil de calcular pero tiene muchas críticas. Es por eso que muchas veces es mejor probar si la distribución de nuestros datos sigue alguna distribución teórica, lo que nos permite hacer otras predicciones con los datos obtenidos. A continuación se describen tres distribuciones de probabilidad, de datos discretos, que más se utilizan para explicar datos que siguen los tres tipos básicos de dispersión que se pueden encontrar en los seres vivos. La distribución de Poisson que explica apropiadamente la dispersión aleatoria, la Binomial para datos de dispersión uniforme y la Binomial Negativa que explica muchas de las situaciones de dispersión agrupada. Distribuciones estadísticas hay muchas y los datos podrían ajustarse bien a varias o también puede que se ajusten mejor a algunas distribuciones con las que no se están comparando los datos.

Comprando con una distribución teórica de probabilidad

Para comparar con otras distribuciones teóricas primero debemos conocer algunos parámetros que describen la distribución de nuestros datos, como el promedio, la desviación estándar el tamaño de muestra, etc. Para comparar con algunas distribuciones se deben estimar otros parámetros como por ejemplo k (número de eventos exitosos en n intentos) o la probabilidad p (probabilidad de un evento exitoso).

```
promedio<-mean(datos$individuos) # calculo promedio
desvest<-sd(datos$individuos) # calculo desviación estándar
mi.var<-var(datos$individuos) # calculo variancia
mi.n<-length(datos$individuos) # calculo tamaño de muestra
mi.max<-max(datos$individuos) # número máximo de individuos por parcela
```

Para comparar con una distribución teórica estadística se deben estimar los valores esperados según esa distribución teórica. Hay paquetes dentro del programa R que usan distintos métodos para comparar un conjunto de datos con distribución teórica. El paquete *fitdistrplus* sirve para ese objetivo. Vamos a comparar el conjunto de datos de palmas con las tres distribuciones teóricas mencionadas.

```
library("fitdistrplus")
fp <- fitdist(datos$individuos, "pois") # mide el ajuste de los datos a una distribución de Poisson

## $start.arg
## $start.arg$lambda
## [1] 7.35
##
##
## $fix.arg
## NULL

fbn<-fitdist(datos$individuos, "nbinom") # mide el ajuste de los datos a una distribución Binomial Negativa

## $start.arg
## $start.arg$size
## [1] 5.838692
##
## $start.arg$mu
## [1] 7.35
##
##
## $fix.arg
## NULL

# Para comparar con una distribución binomial se necesitan dos parámetros extra, el valor k y el valor p
est.k<-ceiling(max(c(promedio^2/(promedio-mi.var), max(datos$individuos))))
est.k

## [1] 23
```

```
est.p<-promedio/est.k  
est.p
```

```
## [1] 0.3195652
```

```
fb<-fitdist(datos$individuos, "binom", fix.arg=list(size=est.k), start=list(prob=est.p))
```

```
## $start.arg  
## $start.arg$prob  
## [1] 0.3195652  
##  
##  
## $fix.arg  
## $fix.arg$size  
## [1] 23
```

```
summary(fp)
```

```
## Fitting of the distribution ' pois ' by maximum likelihood  
## Parameters :  
##      estimate Std. Error  
## lambda      7.35  0.3031089  
## Loglikelihood: -236.9307   AIC:  475.8614   BIC:  478.2434
```

```
summary(fbn)
```

```
## Fitting of the distribution ' nbinom ' by maximum likelihood  
## Parameters :  
##      estimate Std. Error  
## size 5.912031  1.7375265  
## mu   7.349707  0.4539504  
## Loglikelihood: -220.2498   AIC:  444.4996   BIC:  449.2637  
## Correlation matrix:  
##           size      mu  
## size 1.0000000000 0.0001050407  
## mu   0.0001050407 1.0000000000
```

```
summary(fb)
```

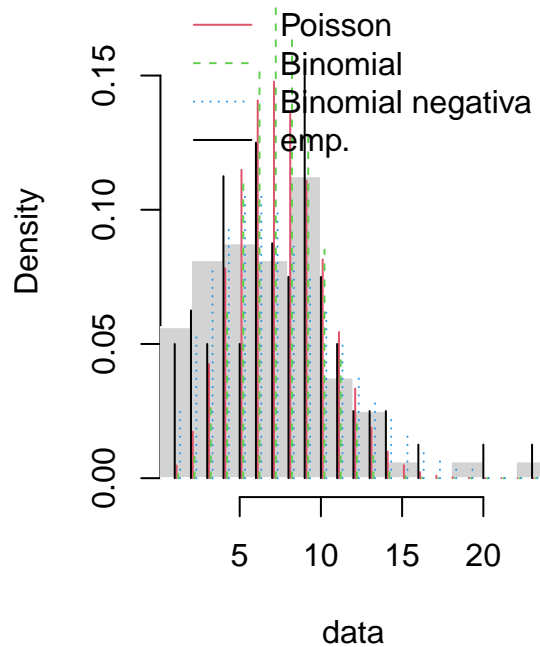
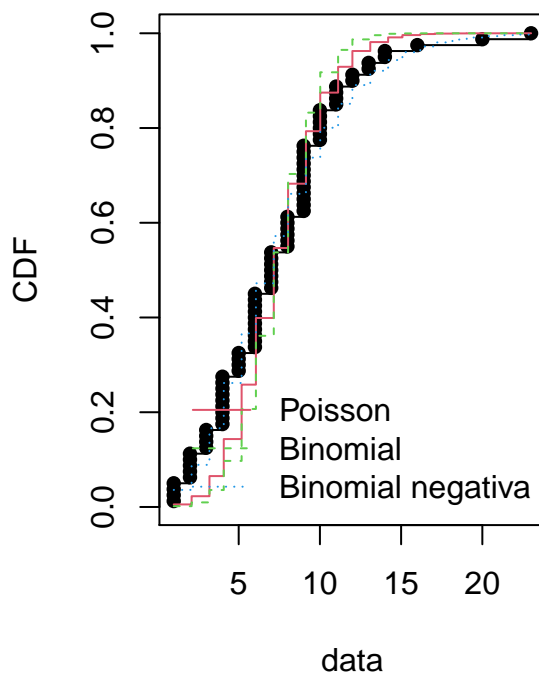
```
## Fitting of the distribution ' binom ' by maximum likelihood  
## Parameters :  
##      estimate Std. Error  
## prob 0.3195652 0.01087078  
## Fixed parameters:  
##      value  
## size    23  
## Loglikelihood: -271.7249   AIC:  545.4499   BIC:  547.8319
```

En los “summary” el paquete hace una estimación de los parámetros usados en cada ajuste. Además presenta los valores de versimilitud y criterios de información que nos ayudan a decidir qué distribución se ajusta mejor a los datos.+


```
library("fitdistrplus")
par(mfrow = c(1, 2))
# Gráfico de Función acumulativa de la distribución.
cdfcomp(list(fp, fb, fbn), legendtext=c("Poisson", "Binomial",
                                         "Binomial negativa"))

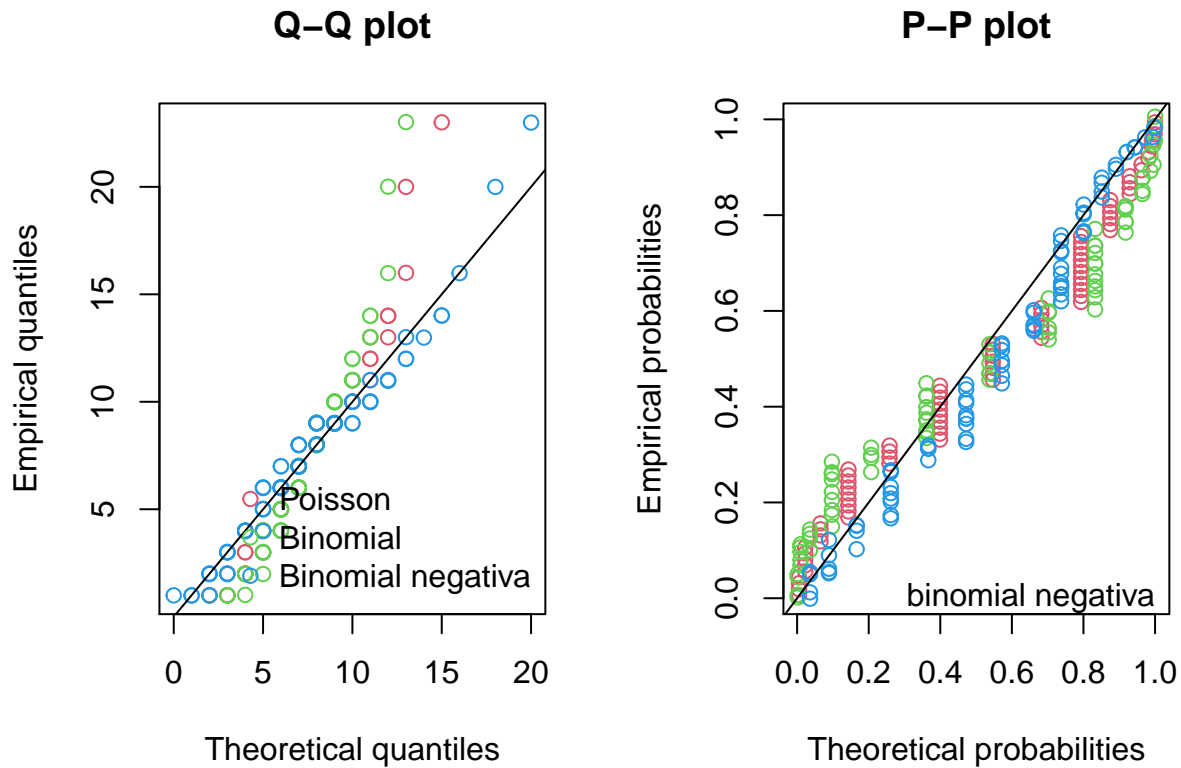
# Gráfico de Función de densidad.
denscomp(list(fp, fb, fbn), legendtext=c("Poisson", "Binomial",
                                         "Binomial negativa"))
```

Empirical and theoretical CDFs Histogram and theoretical densiti



```
# Gráfico de quantiles teóricos.
qqcomp(list(fp, fb, fbn), legendtext=c("Poisson", "Binomial",
                                         "Binomial negativa"))

# Gráfico de probabilidades teóricas.
ppcomp(list(fp, fb, fbn), legendtext=c("Poisson", "Normal", "
                                         binomial negativa"))
```



```
gofstat(list(fp, fb, fbn), fitnames=c("Poisson", "Binomial",
                                       "Binomial negativa"))
```

```
## Chi-squared statistic: 38.84449 116.8265 7.79779
## Degree of freedom of the Chi-squared distribution: 5 5 4
## Chi-squared p-value: 2.552285e-07 1.474605e-23 0.09927263
## the p-value may be wrong with some theoretical counts < 5
## Chi-squared table:
##      obscounts theo Poisson theo Binomial theo Binomial negativa
## <= 2          9      1.817830      0.7711969      7.113906
## <= 4         13      9.653219      7.0070320     13.819536
## <= 6         14     20.446125     21.1272888     16.844354
## <= 8         13     22.679046     27.3324936     15.107132
## <= 9         12      8.868467     10.3623832      6.110575
## <= 11        10     10.873749     10.5950261      9.098437
## > 11         9      5.661565      2.8045793     11.906060
##
## Goodness-of-fit criteria
##
##               Poisson Binomial Binomial negativa
## Akaike's Information Criterion 475.8614 545.4499      444.4996
## Bayesian Information Criterion 478.2434 547.8319      449.2637
```

La función `geofstat`, presenta categorías de frecuencias, valores observados y teóricos esperados para cada distribución. Además de valores de criterios de información. Para seleccionar usando el Criterio de Información de Akaike (AIC) se debe buscar cuál distribución tiene el valor de AIC más bajo. En este ejemplo, sería

la distribución binomial negativa, lo que quiere decir que las palmas presentan un patrón de distribución agregado.

Observando los resultados es evidente que la distribución de nuestros datos difiere menos de la distribución binomial negativa lo que sugiere que las palmas tienen una distribución agrupada. Sin embargo, existen otros tipos de distribuciones que se utilizan para modelar distribuciones agrupadas, por ejemplo, la distribución de Thomas, series de distribución logarítmica, la distribución Pólya-Aeppli y las distribuciones de Neyman tipo A, B y C (Coly et al. 2016).

Referencias

Coly, S., Yao, A. F., Abrial, D., & Charras-Garrido, M. (2016). Distributions to model overdispersed count data. *Journal de la Société Française de Statistique*, 157(2), 39-64.

Krebs, C. J. 1999. *Ecological Methodology*. Segunda Edición. Addison-Welley Educational Publishers, Menlo Park, Ca.