

SeqSam, un programa para la elaboración de planes de muestreo secuencial

Fernando Casanoves¹
Julio A. Di Rienzo²

RESUMEN. Se presenta la descripción de un programa para realizar planes de muestreo secuencial bajo diferentes tipos de distribuciones estadísticas aplicables a conteos de insectos. El programa permite además hacer pruebas de bondad de ajuste a las diferentes distribuciones con el fin de seleccionar la mejor para modelar los datos de conteo. Para facilitar el desarrollo de los planes de muestreo secuencial, se brindan también las curvas de potencia y los intervalos de confianza para la continuidad del muestreo en función del tamaño de muestra. La aplicación está disponible en <http://agro.uncor.edu/~estad/SeqSam.exe>.

Palabras clave: distribución espacial, Ley de Taylor, muestreo de insectos, nivel de daños económicos, umbral económico de daños.

ABSTRACT. *SeqSam, a program for elaborating sequential sampling plans.* We describe a program for elaborating sequential sampling plans under different types of statistical distributions concerning insect counts. The program enables goodness of fit tests of the different distributions in order to select the best one for modeling data. To render the development of sequential sampling plans easier, the program offers power curves and confidence intervals for sampling continuity depending on sample size. The program is free, available at <http://agro.uncor.edu/~estad/SeqSam.exe>.

Keywords: economic damage levels, economic damage thresholds, insect sampling, spatial distribution, Taylor's Law.

Introducción

En el muestreo de insectos para evaluar niveles de daño, en general el interés consiste en contar el número de insectos por unidad muestral o el número de unidades de observación que presentan insectos (o daño) dentro de una unidad muestral. Por ejemplo, si la unidad muestral es una planta, entonces se pueden contar todos los insectos que hay en la planta o contar el número de hojas atacadas dentro de ella. Un plan de muestreo requiere definir el número de unidades muestrales por seleccionar. Esta cantidad depende del patrón de distribución de la plaga y de la precisión de las estimaciones que se quieren realizar. Si el nivel de la plaga es muy bajo o muy alto, usualmente una muestra pequeña permite establecer si hay que tomar medidas correctivas o no; por lo tanto, desde un punto de vista práctico no es conveniente partir de una muestra muy grande cuando con una muestra menor se pueden obtener resultados igualmente confiables. Esta es la idea que subyace el muestreo conocido como “secuencial”.

El *muestreo secuencial* fue introducido por Wald en los años 40 (Binns et ál. 2000) y extendido por Iwao (1975). Consiste en tomar una muestra inicial de tamaño mínimo e ir aumentando el tamaño total de la muestra según un plan de muestreo fijado, contando el número de eventos acumulados (conteo de insectos, hojas con insectos, hojas atacadas, etc.) observados en el muestreo secuencial. En cada paso de la secuencia puede ocurrir que se requiera seguir incrementando el tamaño muestral total o que se tome una decisión basada en el nivel poblacional estimado hasta el momento. Para saber si el muestreo continúa o se detiene, el conteo acumulado se compara con dos valores que delimitan una región de incertidumbre. Si el conteo acumulado está por debajo del límite inferior de esta región, entonces se concluye que el nivel de la plaga se encuentra por debajo de su densidad crítica (dc). Por el contrario, si el conteo acumulado está por encima del límite superior se concluye que la plaga supera la dc . En cualquiera de los dos casos

¹ Unidad de Bioestadística, CATIE, Sede Central, Turrialba 7170, Costa Rica. casanoves@catie.ac.cr

² Estadística y Biometría, Universidad Nacional de Córdoba, Argentina. dirienzo@agro.uncor.edu

anteriores el muestreo llega a su fin. Si, en cambio, el conteo acumulado está dentro de la región de incertidumbre, el tamaño muestral se sigue incrementando hasta que el conteo acumulado cae fuera de la región de incertidumbre o hasta alcanzar un tamaño muestral máximo establecido *a priori*. Para establecer los límites de la región de incertidumbre se necesita conocer la *dc* de la plaga (como media del número de conteos por unidad muestral) y contar con un modelo para la distribución de frecuencias de los conteos. Se supone que la población muestreada permanece constante durante el proceso de muestreo.

El programa descrito en este trabajo calcula *curvas de decisión* y *curvas de potencia* para un muestreo secuencial a partir de la especificación de la *dc* de la población y el modelo estadístico para la distribución de frecuencia del conteo de insectos por unidad muestral. Asimismo, calcula pruebas de bondad de ajuste a modelos de distribución teóricos para conteos a partir de una muestra piloto.

Fundamentación

Iwao (1975) propuso la construcción de las curvas de decisión basándose en una aproximación normal, esto es, límites de la región de incertidumbre (LRI) dados por:

$$dc \pm z_{1-\alpha/2} \sqrt{\sigma^2/N}, \quad [1]$$

donde *dc* es la densidad crítica de la población, $z_{1-\alpha/2}$ el cuantil $(1-\alpha/2)$ de una distribución normal estándar, σ^2 la varianza de los conteos y *N* el tamaño muestral acumulado.

Dada la *dc*, expresada como la media del número de conteos por unidad muestral, el parámetro que debe ser estimado para completar la especificación del modelo normal es la varianza. Bajo un modelo normal clásico, la varianza es independiente de la media y podría calcularse a partir de cualquier muestra piloto. Sin embargo, es bien sabido que la distribución de conteos no sigue la ley normal sino otras leyes en las que generalmente la varianza depende de la media (ejemplos de estos otros modelos son Poisson, binomial negativa, binomial y beta binomial, entre otros). En los casos en que la varianza depende de la media, no toda muestra piloto sirve para estimar la varianza. Cuando la media de conteos es muy distinta a la *dc*, la estimación de la varianza no es correcta. Una solución para este problema de estimación consiste en encontrar una función que vincule la varianza con la media. Taylor (1984) propuso el modelo de potencia para relacionar la varianza con la media según la siguiente expresión: $\sigma^2 = f(\mu) = \alpha\mu^\beta$, donde μ es la media poblacional (promedio de los conteos), y α y β parámetros que deben estimarse. La estimación se realiza mediante un análisis de regresión lineal suponiendo el modelo: $\hat{y} = \alpha^* + \beta x$, donde

$\hat{y} = \ln(\sigma^2)$, $\alpha^* = \ln(\alpha)$ y $x = \ln(\mu)$. La Ley de Taylor ha mostrado ser de gran aplicabilidad bajo distintos modelos de distribución de conteos. Una vez estimados los parámetros bajo esta ley, la varianza de los conteos se calcula tomando la *dc* como argumento de la función. Luego, se aplica la Expresión 1 para obtener los LRI para distintos tamaños muestrales. Esta forma de abordar la construcción de los LRI demanda la disponibilidad de al menos tres muestras donde la media varíe alrededor de la *dc* para poder estimar mediante regresión lineal los parámetros del modelo de Taylor. Por otra parte, la construcción de los LRI utilizando la Expresión 1 produce límites simétricos alrededor de la *dc*, lo cual puede no ser una aproximación apropiada para muchas distribuciones de conteos.

La alternativa para la construcción de estos límites es suponer algún modelo teórico plausible para los conteos.

Naranjo y Hutchinson (1997) desarrollaron un programa para validación de planes de muestreo secuencial, que permite validar la distribución y obtener curvas de decisión basadas en el modelo de Taylor. La diferencia de esta propuesta con la que se presenta aquí es que SeqSam permite obtener planes de muestreo secuencial basados en la varianza de las distribuciones, en vez de usar estimaciones obtenidas a través del modelo de Taylor.

En la aplicación del programa descrito en este trabajo se han implementado cuatro modelos de conteo: dos para modelar el número de éxitos (insectos) por unidad muestral (Poisson y binomial negativa) y dos para modelar el número de unidades observacionales con al menos un éxito dentro de una unidad muestral que contiene *n* unidades observacionales (binomial y beta binomial). Cada uno de estos modelos tiene parámetros que los caracterizan, pero existen diversas formas de parametrizarlos. En esta presentación se ha seguido la siguiente parametrización, donde *m* es la media poblacional del número de conteos; *Binomial Negativa(m,k)*, donde *k* es un parámetro que modula la agregación (ocurrencia de pocas unidades muestrales con gran cantidad de éxitos y muchas con pocos o ninguno) y *Binomial(m,n)*, donde *n* es el número de unidades de observación en cada unidad muestral (conglomerado). El parámetro *n* se supone mayor o igual que *m*. Cuando *n* es igual a 1, la distribución binomial se conoce como “distribución Bernoulli” y el parámetro *m* corresponde a la probabilidad de éxito. Por último, la distribución *Beta Binomial(m,n,r)* tiene los mismos parámetros que la Binomial, excepto que se agrega *r*, el coeficiente de *correlación intraclase*, que modela la falta de independencia en la ocurrencia de éxitos dentro de un conglomerado (contagio). En el Recuadro 1 se presenta una descripción de las funciones de probabilidad de los modelos y su parametrización.

Recuadro 1. Distribuciones estadísticas

Distribución binomial

Las variables binomiales con parámetros m, n se denotan como: $X \sim \text{Bin}(m, n)$ y su función de probabilidad esta dada por:

$$p(X = x | m, n) = C(n, x) \left(\frac{m}{n}\right)^x \left(1 - \frac{m}{n}\right)^{1-x}, \quad x = 0, 1, \dots, n.$$

donde $C(n, x) = \frac{n!}{x!(n-x)!}$ y representa el número de formas posibles de tomar x elementos de un grupo de n .

Distribución binomial negativa

A las variables binomiales negativas con parámetros m y k se las denota como

$X \sim \text{BinNeg}(m, k)$ y su función de probabilidad esta dada por:

$$P(X = x | m, k) = \frac{\Gamma(k+x)}{\Gamma(x+1) * \Gamma(k) * \left(\frac{m}{k+1}\right)^k * \left(\frac{(m/k)}{(m/k+1)}\right)^x}, \quad x = 0, 1, 2, \dots$$

Distribución Poisson

Se denota a una Poisson con parámetro m como: $X \sim \text{Poisson}(m)$ y su función de probabilidad esta dada por:

$$p(X = x | m) = \frac{m^x e^{-m}}{x!}, \quad x = 0, 1, 2, \dots$$

Distribución beta binomial

A las variables beta binomiales con parámetros m, n y r se las denotará como

$X \sim \text{BetaBin}(m, n, r)$ y su función de probabilidad esta dada por:

$$p(X = x | n, m, r) = C(n, x) \frac{\Gamma(a+b) \Gamma(a+x) \Gamma(b+n-x)}{\Gamma(a) \Gamma(b) \Gamma(a+b+n)}, \quad x = 0, 1, 2, \dots, n.$$

donde $a = \frac{m(1-r)}{n(r)}$ y $b = \frac{(n-m)1-r}{n(r)}$

Cuando los modelos elegidos son $\text{Poisson}(m)$ o $\text{Binomial}(m, n)$ (n es siempre un parámetro conocido), para la construcción de los LRI, m se iguala a la dc , pero cuando los conteos siguen las distribuciones $\text{Binomial Negativa}(m, k)$ o $\text{Beta Binomial}(m, n, r)$, los parámetros k y r deben ser provistos ya sea a partir de valores previamente establecidos o de su estimación a partir de una muestra piloto.

Una vez que el modelo ha sido seleccionado y sus parámetros especificados, los LRI se calculan como los correspondientes cuantiles $Q_{(\alpha/2)}$ (límite inferior) y $Q_{(1-\alpha/2)}$ (límite superior) de la distribución de los conteos acumulados. Se puede suponer que los conteos acumulados

son la suma de N variables aleatorias independientes idénticamente distribuidas según la especificación del modelo para los conteos seleccionados. En algunos modelos esta distribución es conocida y en otros no. En estos casos se pueden usar aproximaciones basadas en el método de simulación Monte Carlo u otras aproximaciones basadas en supuestos adicionales, como el uso de la aproximación normal.

En un muestreo secuencial debe establecerse el tamaño de la muestra inicial y el tamaño muestral máximo que se está dispuesto a tomar en el caso de que no se pueda parar en un paso anterior de la secuencia de

muestreo. Para establecer estos tamaños muestrales es necesario introducir el concepto de *potencia*. La potencia de una prueba estadística se calcula como la probabilidad de que la prueba detecte que la hipótesis nula es falsa. En el contexto del muestreo secuencial, la hipótesis que se somete a prueba es que la media del número de conteos en la población es igual a la dc . En cada paso de la secuencia de muestreo esta hipótesis es sometida a prueba utilizando como estadístico el conteo total acumulado. Si este cae en la región de incertidumbre, la hipótesis no puede desecharse ni por exceso ni por defecto, y por lo tanto el muestreo debe continuar para obtener una mayor cantidad de información. En caso contrario, se dice que la densidad de la población está por encima o por debajo de la dc y se procede en consecuencia.

Un criterio estadístico para fijar el tamaño muestral inicial es buscar aquel valor de N para el cual la potencia de la prueba es mayor o igual a un valor razonable (por ejemplo del 80%). Para calcular la potencia hace falta especificar la discrepancia (por exceso o defecto) entre la densidad poblacional y la dc . El N máximo se escoge de tal manera que la potencia sea, por ejemplo, mayor al 80-90%. La potencia se incrementa con N y con la magnitud de la discrepancia que se quiere detectar. Por lo tanto si el investigador quiere detectar una diferencia entre la densidad poblacional y la dc del orden del 1% de la dc , necesitará un tamaño muestral mayor que si quiere detectar una diferencia del orden del 10%. El programa que se describe en este trabajo calcula, para el modelo de conteo seleccionado y una densidad media hipotética (distinta) de la dc , una curva de potencia en función del tamaño muestral, siguiendo la expresión $\pi = p(C \leq Q_{\alpha/2}^N) + p(C \geq Q_{1-\alpha/2}^N)$, donde $Q_{(\alpha/2)}$ y $Q_{(1-\alpha/2)}$ son los cuantiles del conteo total de éxitos (C) para muestras de tamaño N . El tamaño muestral se hace variar entre un valor inicial y uno final, provistos por el usuario. La curva obtenida permite evaluar la pertinencia de los tamaños muestrales propuestos con base en la potencia.

Descripción de la operatoria del programa SeqSam

El programa SeqSam calcula curvas de decisión para un muestreo secuencial basándose en la propuesta de Iwao (1975). Si la presencia de un insecto en una planta (unidad muestral) es el evento de interés, la técnica implementada en esta aplicación permite decidir si el número medio de eventos observados por unidad muestral en la población muestreada supera o está por debajo de una dc establecida previamente. Generalmente, el número de eventos en la muestra estará por debajo o por encima de su valor esperado. Esa variabilidad de los resultados muestrales implica la necesidad de utilizar criterios estadísticos para

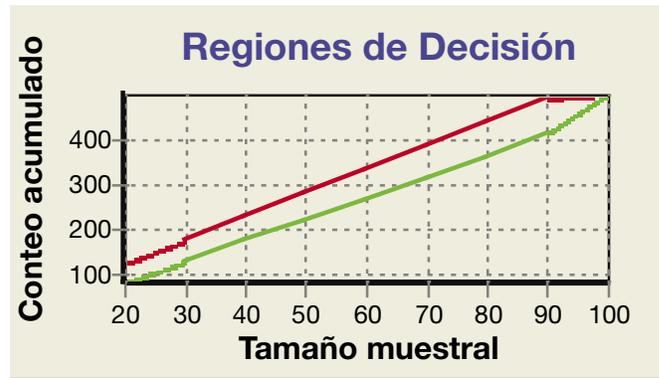


Figura 1. Pantalla de la aplicación con las regiones de decisión.

decidir si el nivel medio de eventos por unidad muestral está por encima o por debajo de la dc preestablecida. La técnica implementada en este programa delimita tres regiones en el plano definido por el tamaño muestral (Eje x) y el número total de eventos en la muestra (Eje y). Sobre este plano, el programa dibuja dos curvas. Debajo de la primera, de color verde, se encuentra la *región de rechazo por defecto* (Figura 1). Un punto en esta región indica que el número promedio de eventos por unidad muestral en la población es menor a la dc . La otra curva, de color rojo, delimita la *región de rechazo por exceso*. Un punto por arriba de esta curva conduce a aceptar que el número promedio de eventos por unidad muestral en la población es mayor que la dc .

La región comprendida entre ambas curvas es la *región de incertidumbre* y un punto en ella no permite tomar una decisión satisfactoria sobre el nivel medio de eventos por unidad muestral. Si para una muestra de tamaño N el número total de eventos (conteo acumulado) en la muestra está en la región de incertidumbre, entonces se agregan unidades muestrales adicionales; en caso contrario, se toman las acciones consistentes con la conclusión sobre el nivel

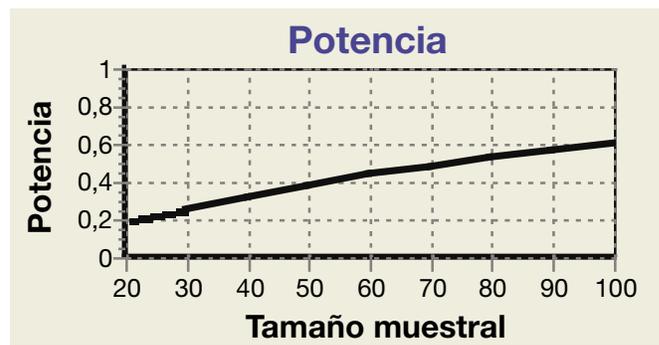


Figura 2. Pantalla de la aplicación mostrando la curva de potencia.

medio de eventos. Si en un plan de muestreo, el primero y los sucesivos conteos acumulados caen en la *región de incertidumbre*, el muestreo sigue, acumulando unidades muestrales, hasta que se alcanza el tamaño muestral máximo (N_{max}). En este caso se decide sobre el número medio de eventos por unidad muestral en la población según si el número total de eventos en la muestra esta por encima o por debajo de su valor esperado suponiendo que el número promedio de eventos en la población es igual a la dc .

Dadas las *regiones de rechazo* y la *región de incertidumbre* el programa puede calcular la probabilidad, para un tamaño muestral dado, de que un punto caiga en las regiones de rechazo si el nivel medio de eventos por unidad muestral es mayor o menor que la dc . Si este cálculo se repite para distintos tamaños muestrales, se obtiene una *curva de potencia* en función del tamaño muestral y sirve para establecer la calidad del plan muestral propuesto.

Construcción de las curvas que delimitan la región de incertidumbre

Para obtener las curvas que delimitan la región de incertidumbre accione el botón rotulado “Obtener Regiones de Decisión” (Figura 3). Debido a que, salvo para

el caso de la distribución de Poisson, la suma de variables aleatorias para los demás modelos de ajuste utilizados no tiene distribución conocida, el programa encuentra los cuantiles de su distribución por simulación Monte Carlo. Para ello, SeqSam genera M muestras (por defecto $M = 500$) de distintos tamaños (N) entre el tamaño muestral mínimo (N_{min}) y N_{max} . El número de puntos en el que se obtienen las M muestras está definido por el campo *puntos*. Luego, para cada tamaño muestral considerado se obtienen M observaciones del número total de conteos por muestra. La región de incertidumbre se delimita por los percentiles muestrales ($alfa/2$) y $(100 - alfa/2)$ estimados a partir de las M observaciones. La región de incertidumbre cubre $(100 - alfa)\%$ de los resultados muestrales cuando la media en la población es igual al dc . Luego, si $alfa = 5\%$, la región de incertidumbre tiene una cobertura del 95%.

Para generar las M muestras es necesario especificar la distribución de los conteos desde donde se obtendrán las muestras. El programa permite seleccionar entre un conjunto común de modelos para conteos: *Binomial(m,n)*, *Poisson(m)*, *BinomialNegativa(m,k)* y *BetaBinomial(m,n,r)* (Figura3). Para todos los modelos, el parámetro m se iguala automáticamente a la dc . Para los modelos de conteo

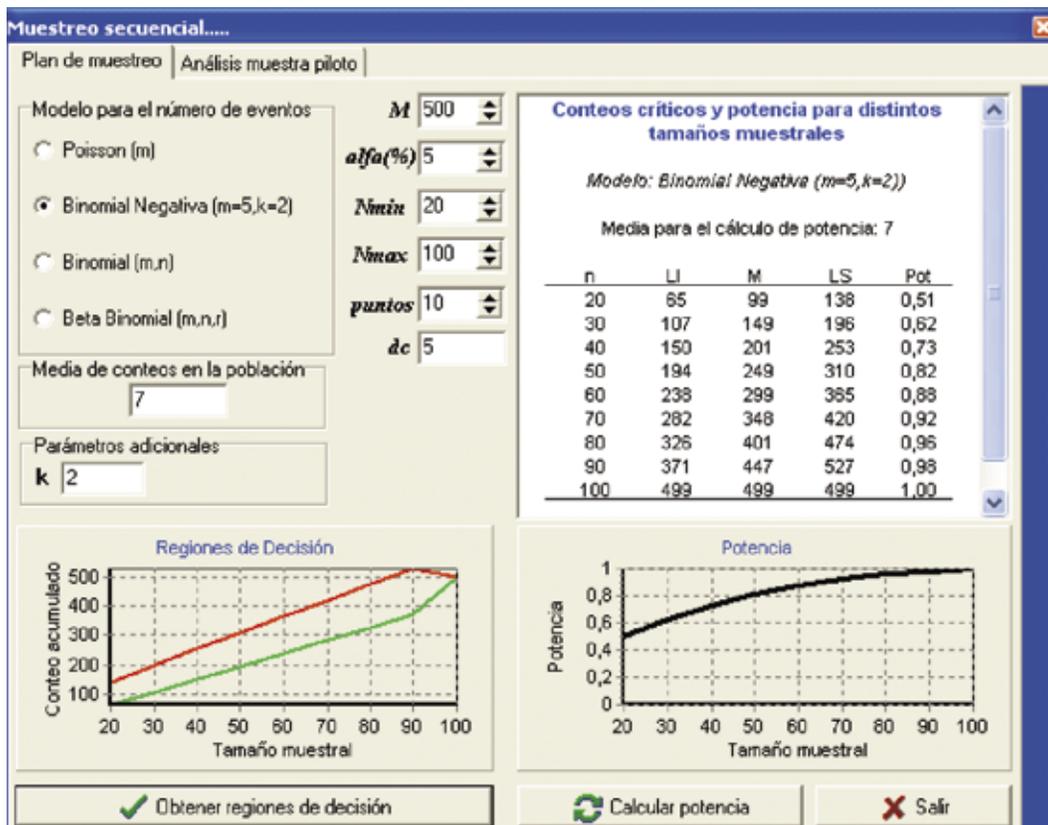


Figura 3. Pantalla de la aplicación con un ejemplo en el que el modelo de conteo es una Binomial Negativa con media $m = 5$ y parámetro de agregación $k = 2$. Los tamaños muestrales varían de 20 a 100 en incrementos de 10 unidades muestrales.

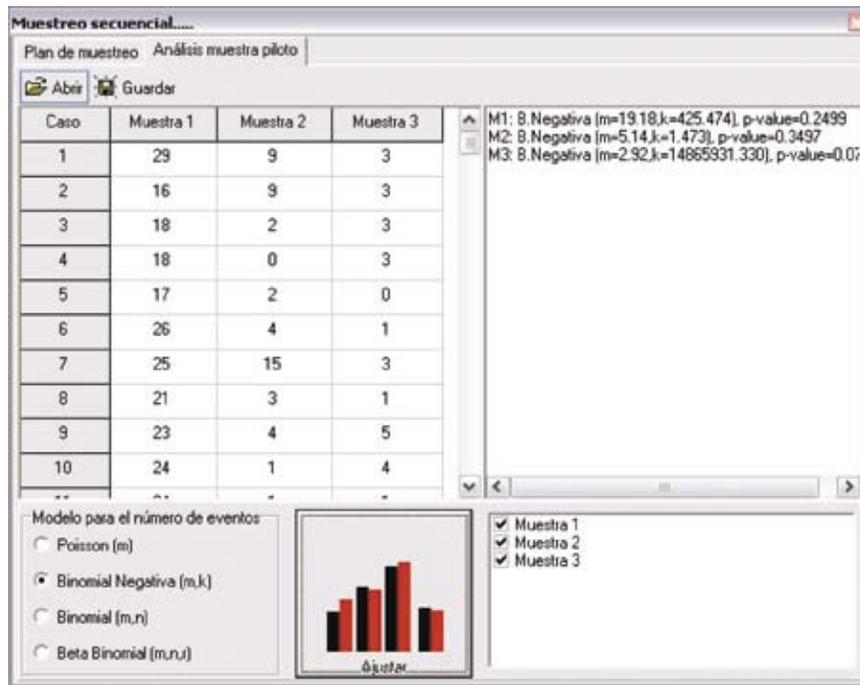


Figura 4. Pantalla que muestra el ajuste de una distribución Binomial Negativa a las muestras 1, 2 y 3. A la derecha se observan los estadísticos de ajuste.

binomial, binomial negativa y beta binomial se deben especificar *parámetros adicionales*. Para la binomial se debe especificar el valor de n (tamaño de conglomerado), para la binomial negativa se debe aportar el valor de k (el parámetro de agregación) y para la beta binomial, el valor de n (tamaño del conglomerado) y r (la correlación intraclase). El tamaño del conglomerado para la binomial y la beta binomial no es un problema porque se conoce (es una decisión del plan de muestreo), pero los parámetros r y k usualmente deben estimarse a partir de un muestreo piloto. El análisis de una muestra piloto se abordará más adelante.

Construcción de la curva de potencia

Una preocupación natural en el diseño de un plan de muestreo consiste en determinar el tamaño de la muestra. En un muestreo secuencial hay que decidir los tamaños muestrales N_{min} y N_{max} . Para establecer cuáles son los valores razonables para estas especificaciones del plan muestral, se puede utilizar el concepto estadístico de potencia. Para obtener la *curva de potencia*, el usuario debe accionar el botón *Calcular potencia* (Figura 3). La curva de potencia (Figura 2) depende del verdadero valor de la media de eventos en la población; este valor nunca se conoce pero el investigador puede proponer valores alternativos de la *Discrepancia con la dc* que tengan algún sentido práctico para obtener la curva de potencia.

En la Figura 3 se presenta la pantalla de diálogo con el usuario. El modelo elegido es una Binomial Negativa con media $m = 5$ y parámetro de agregación $k = 2$, y tamaños muestrales que van de 20 a 100 en incrementos de 10. Con estos tamaños muestrales se obtiene una potencia inicial mayor a 0,40 y una potencia final cercana a 1. Estas potencias se obtuvieron suponiendo que la media poblacional no es 5 sino 7. La tabla de *conteos acumulados críticos* indica que para una muestra de tamaño $n = 90$, un número total de eventos en la muestra, inferior a 371 o mayor a 527, indica que la media poblacional no es 5. La curva de potencia para este ejemplo permite inferir que para alcanzar una potencia de 0,80 (valor mínimo recomendado en estudios biológicos) se requiere de un tamaño muestral de 50; por lo tanto, este tamaño debería ser el mínimo en el plan de muestreo secuencial.

Dado que tanto la estimación de la *región de incertidumbre* como la curva de potencia se obtienen por simulación Monte Carlo, estas pueden variar levemente cada vez que se corre el programa, aunque estas variaciones no tienen demasiada importancia práctica. De todos modos, las variaciones entre corridas pueden controlarse, si fuera necesario, aumentando M .

Si el usuario acciona el botón derecho del ratón sobre la tabla de *Conteos críticos* o sobre los gráficos de las regiones de decisión o el de potencia, podrá copiarlos para llevarlos a otras aplicaciones de Windows.

Recuadro 2. Ejemplo de aplicación del programa

A partir de un muestreo preliminar se tomaron 50 datos de conteo de insectos (Cuadro 1) sobre 5 plantas (unidad muestral). Debido a la naturaleza de la toma de datos se espera que los conteos tengan distribución Poisson.

Cuadro 1. Valores de conteos de insectos obtenidos en un muestreo piloto con $N = 50$

5	5	3	5	5
9	3	4	7	6
5	9	9	6	6
3	4	4	9	6
7	3	4	4	6
8	2	7	3	6
7	5	6	2	1
5	6	3	5	7
3	4	3	3	3
5	1	4	4	3

Utilizando la opción *Análisis de muestra piloto* de SeqSam se ajustaron las cuatro distribuciones y se obtuvieron los siguientes resultados:

- Poisson ($m = 4,85$), $p = 0,4916$
- B.Negativa ($m = 4,85$, $k = 18358183,589$), $p = 0,3326$
- Binomial ($m = 4,85$, $n = 9$), $p = 0,0076$
- B.Binomial ($m = 4,85$, $n = 10$, $r = 0,072$), $p = 0,0013$

Estos resultados indican que la distribución más probable para estos conteos es la Poisson, debido a que tiene el valor p más alto ($p = 0,4916$). Se puede decir entonces que los conteos de la muestra piloto siguen una distribución Poisson con media de conteos en la población = 4,85.

Luego, si la dc para el cultivo es de 6, se puede ahora obtener el plan de muestreo secuencial a partir de una distribución Poisson con $m = 6$. El tamaño mínimo para comenzar el muestreo secuencial debe ser de 40 si se requiere una potencia superior a 0,8. (Figura 5).

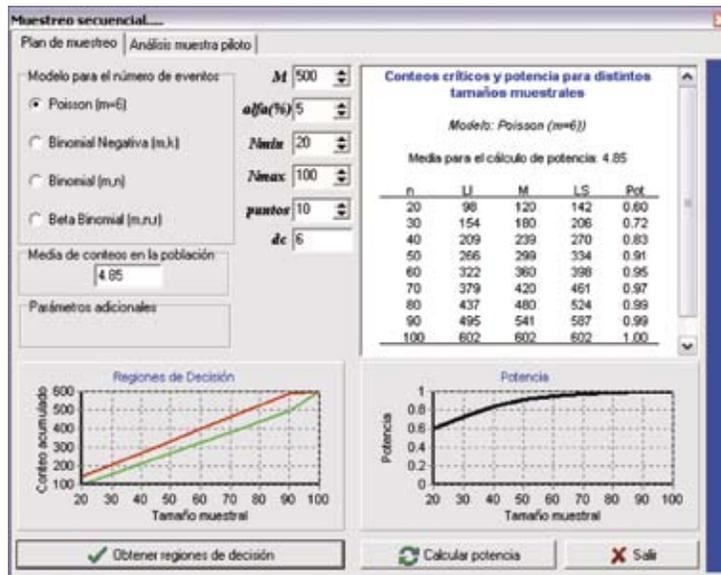


Figura 5. Pantalla de la aplicación con un ejemplo en el que el modelo de conteo es una Poisson con media $m = 6$. Los tamaños muestrales varían de 20 a 100 en incrementos de a 10 unidades muestrales.

Luego, si con una muestra de 40 observaciones el conteo acumulado de insectos es menor a 209, el muestreo finaliza concluyendo que la plaga esta por debajo de la dc ; si el conteo acumulado supera 270, el muestreo finaliza concluyendo que la plaga está por encima de la dc ; en caso contrario, es decir, conteos acumulados entre 209 y 270, se concluye que hay que continuar el muestreo secuencial. En este caso, se toman otras 10 unidades muestrales para completar un total de 50, y con el conteo acumulado se repite el procedimiento anterior para detener o continuar con el muestreo secuencial.

Análisis de muestra piloto

El programa SeqSam permite analizar una muestra piloto con el doble propósito de evaluar la bondad de ajuste de alguno de los modelos teóricos utilizados en el programa y estimar los parámetros k y r de las distribuciones binomial negativa y beta binomial, respectivamente. Para utilizar esta herramienta, el usuario debe seleccionar la solapa *Análisis muestra piloto* (Figura 4). En esta solapa el usuario abre un archivo con los conteos de una o más muestras piloto, utilizando el botón *Abrir*. El programa acepta tres formatos de datos: texto (.txt), Excel (.xls) o un formato propio del programa (.sq). Si el usuario decide guardar los datos en el formato propio de SeqSam, debe accionar el botón *Guardar*. Se supone que los datos de una muestra piloto son conteos y están en una columna del archivo de datos. Si existen varias columnas, el programa supone que cada una de ellas es una muestra piloto distinta y las analiza separada y consecutivamente, designándolas en la salida como $M1$, $M2$,... y así sucesivamente. Si el archivo de datos contiene columnas que no se desea analizar, se pueden seleccionar/deseleccionar estas columnas haciendo doble clic sobre el encabezamiento de las mismas, o bien, accionando el botón derecho del ratón sobre la grilla que contiene los datos es posible seleccionar/deseleccionar todas las columnas.

Una vez leídos los datos, se habilita la selección del modelo que se quiere ajustar. Para los modelos Binomial y Beta Binomial se debe especificar el tamaño del conglomerado. Si este dato no se especifica o se subespecifica, el programa asumirá que su valor es el máximo conteo en la muestra. Esta suposición automática del tamaño del conglomerado suele ser poco satisfactoria desde el punto de vista de la estimación, por lo que el usuario debe hacer todo lo posible por proveer este dato.

Una vez seleccionado el modelo, para realizar el ajuste el usuario debe accionar el botón *Ajustar*. Como respuesta, el programa presenta una salida que muestra el modelo ajustado con los parámetros estimados y el valor p para la prueba de hipótesis de que el modelo propuesto ajusta a los datos. Para una muestra en particular, puede ocurrir que más de un modelo sea apropiado para ajustar la distribución de frecuencias observada (valor $p > 0,05$). Como regla práctica, excepto que haya un criterio de elección que no dependa de los datos observados, se aceptará como el modelo más plausible aquel cuyo valor p sea mayor. En el caso de múltiples muestras piloto obtenidas sobre la misma población biológica bajo distintas condiciones, debería considerarse la conveniencia de utilizar un modelo común para todas ellas. Un ejemplo de aplicación del programa es desarrollado en el Recuadro 2.

Agradecimientos

Los autores agradecen el financiamiento parcial recibido por USDA FAS CS-31 (Purdue University) para el proyecto Using Geographic Information Systems to Enhance Phytosanitary Conditions and Trade Capacity of Small Growers of Tropical Ornamental Crops.

Literatura citada

- Binns, MR; Nyrop, JP; Van der Werf, W. 2000. Sampling and monitoring in crop protection: The theoretical basis for developing practical decision guides. Reino Unido, CAB International. 284 p.
- Iwao, S. 1975. A new method of sequential sampling to classify populations relative to a critical density. *Researches on Population Ecology* 16:281-288.
- Naranjo, SE; Hutchison, WD. 1997. Validation of arthropod sampling plans using a resampling approach: software and analysis. *American Entomologist* 43(1): 48-57.
- Taylor LR. 1984. Assessing and interpreting the spatial distributions of insect populations. *Annual Review of Entomology* 29: 321-359.