

# MÉTODO SIMPLE PARA TAMAÑO ÓPTIMO DE MUESTRA APLICABLE A PARASITISMO, INFESTACIÓN Y OTRAS VARIABLES DE DOS CATEGORÍAS<sup>1</sup>

Juan F. Barrera y Javier Valle  
El Colegio de la Frontera Sur (ECOSUR), Carretera Antiguo Aeropuerto km 2.5,  
Tapachula, Chiapas, México.

Palabras clave: Muestreo binomial, tamaño de muestra, proporción muestral.

## Introducción

Determinar el tamaño óptimo de muestra es una de las tareas más importantes de cualquier investigación que requiera muestrear una población, no solo para obtener resultados confiables sino también para ahorrar recursos. Un principio que impera en la estimación de un parámetro señala que, para obtener mayor precisión es necesario un tamaño mayor de muestra. Sin embargo, con frecuencia el tamaño de muestra no se determina con relación al número requerido, y la precisión suele ser baja o alta. Esto origina datos de mala calidad en un extremo y costos innecesarios en el otro. En control biológico no es raro encontrar que la población a muestrear se puede clasificar en solo dos categorías: huéspedes parasitados y no parasitados, larvas vivas y muertas, hojas infestadas y no infestadas, etc. En estas situaciones, se tiene interés en saber cual es la proporción, porcentaje o número de individuos en una de las dos categorías. Al muestreo de una población cuyos datos se encuentran en una escala nominal como esta, se le denomina muestreo binomial. El objetivo de este trabajo fue desarrollar un método simple de muestreo para determinar el tamaño óptimo de muestra aplicable a datos que puedan clasificarse en dos categorías.

Planteamiento del problema. La fórmula para estimar el tamaño de muestra ( $n$ ) cuando se quiere estimar la proporción ( $p$ ) de una población con una característica de interés, se puede encontrar en muchos libros de estadística (e.g., Daniel, 1981; 2002):

$$n = (z_{\alpha/2})^2 p q / d^2 \quad \text{Eq. 1}$$

donde  $z$  es la distribución normal estandarizada,  $q = 1 - p$ , y  $d$  es igual a la mitad del intervalo de confianza deseado. Si  $Y$  individuos poseen la categoría de interés y  $N - Y$  son los individuos en la segunda categoría, el valor de  $p$  se obtiene dividiendo  $Y$  entre el número total de individuos ( $N$ ). La Eq. 1 se aplica a muestreos hechos al azar en poblaciones grandes. Cuando no se conoce  $p$  y es impráctico tomar una muestra piloto, el valor máximo de  $n$  se obtiene con la Eq. 1 y  $p = 0.5$  (Daniel, 1981). Por ejemplo, el tamaño de muestra requerido para estimar la proporción de larvas parasitadas en un cultivo con 95% confianza y  $d = 0.05$ , es 384 larvas ( $1.96^2 * 0.5 * 0.5 / 0.05^2$ ). Sin embargo, 384 puede llegar a ser demasiado grande y entonces los costos resultan muy altos.

---

<sup>1</sup> Fuente: Barrera, J. F. & J. Valle. 2004. Método simple para tamaño óptimo de muestra aplicable a parasitismo, infestación y otras variables de dos categorías. En : XXVII Congreso Nacional de Control Biológico. Los Mochis, Sinaloa, México, p. 93-96.

## Materiales y métodos

Propuesta. Como propuesta para abordar este problema, aquí se sugiere tomar una muestra máxima de 390 (valor aproximado a 384, ver arriba) y calcular  $p$  sobre el número acumulado de submuestras de 15 unidades ( $n = 15, 30$  y hasta 390), y  $N$  con valor fijo de 390. Bajo esta situación se incrementa el valor de  $n$  con respecto a  $N$ , y eventualmente  $n/N$  será igual o mayor a 0.05; cuando esto sucede, la corrección por población finita no puede descartarse, y la fórmula más adecuada es (Daniel, 2002):

$$n = N (z_{\alpha/2})^2 p q / [d^2 (N - 1) + (z_{\alpha/2})^2 p q] \quad \text{Eq. 2}$$

Dado que la distribución muestral de  $p$  será aproximadamente normal cuando  $np$  y  $nq$  son mayores que 5, es difícil estimar  $p$  si tiende a estar muy cerca de 0 y 1 (Zar, 1999). Este problema se solucionó estimando  $p$  de los valores de  $np$  y  $nq$  más cercanos y superiores a 5; después, se calculó el intervalo de confianza (IC95%) con:

$$p \pm z_{\alpha/2} * \text{raíz cuadrada de } p q / n \quad \text{Eq. 3}$$

Los valores de  $p$  cercanos a 0 y 1 fueron 0.07 y 0.93, respectivamente, y el IC95% fue de 0.0560 para ambos casos. Así, el tamaño de muestra para estimar  $p$  entre 0.0140 y 0.1260 y entre 0.8740 y 0.9860, fue de 120, mientras que  $n$  para  $p$  entre 0.1261 y 0.8739 fue estimado con la Eq. 2. La Fig. 1 presenta los valores resultantes. Esta figura se usa tomando muestras consecutivas de 15 unidades y determinando la proporción ( $p$ ) de individuos con la característica de interés sobre el número acumulado de  $n$ . Al momento que la relación entre  $n$  y  $p$  sobrepasa la línea presentada en la figura, se deja de muestrear, indicando que el tamaño de muestra observado es suficiente para estimar  $p$  con 95% de confianza y  $d = 0.05$ .

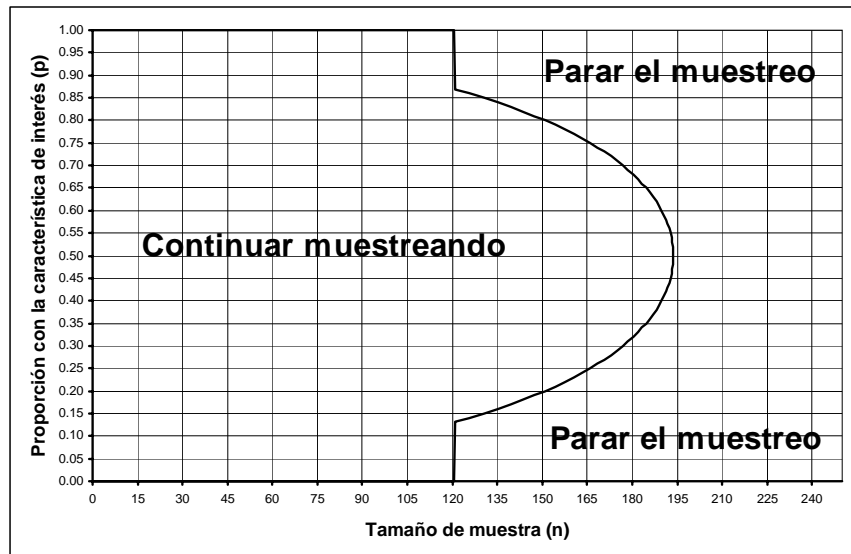


Fig. 1. Relación entre tamaño de muestra ( $n$ ) y proporción ( $p$ ) de la población con la característica de interés;  $p$  se estimó con Eq. 2, y  $z = 1.96$ ,  $d = 0.05$  y  $N = 390$ . La  $n$  para  $p$  entre 0.0140 y 0.1260 y entre 0.8740 y 0.9860, fue de 120.

Los datos. La bondad de este método fue validada con datos de poblaciones de insectos asociados al café en Soconusco y Sierra de Chiapas. Los datos fueron la

proporción de: 1) frutos de café con broca (*Hypothenemus hampei*) y presencia de parasitoides (*Cephalonomia stephanoderis*); 2) ramas con frutos perforados por broca; 3) ramas con frutos y hojas dañadas por chacuatete (*Idiarthron subquadratum*) y 4) ramas con hojas afectadas por minador (*Leucoptera coffeella*). Los datos se analizaron con regresión lineal entre los valores de  $p$  para  $N = 390$  ( $n$  máxima, Eq. 1) con respecto a los valores de  $p$  para  $n$  ( $n$  óptima, Eq. 2 y Fig. 1).

## Resultados y Discusión

La Fig. 2 presenta la relación entre  $p$  de  $N$  y  $p$  de  $n$  para 29 pares de datos de poblaciones de cuatro especies de insectos asociados al café. Gráficamente se puede observar que el ajuste fue bastante bueno. El valor del coeficiente de correlación fue 0.9857. Esto significa que puede ser muy confiable usar la Fig. 1 bajo el procedimiento señalado para determinar el tamaño de muestra cuando se recurre al muestreo binomial. Este método puede ayudar a los practicantes del control biológico a determinar un tamaño de muestra apropiado para estimar, por ejemplo parasitismo, con la posibilidad de evitar gastos elevados como consecuencia de analizar muestras demasiado grandes. Es importante mencionar que esta metodología supone que la probabilidad de observar la característica de interés permanece constante, lo cual puede no ser cierto para algunos casos. En conclusión, esta metodología se considera adecuada para determinar tamaño de muestra en variables con dos categorías cuya muestra preliminar proviene de una población infinita.

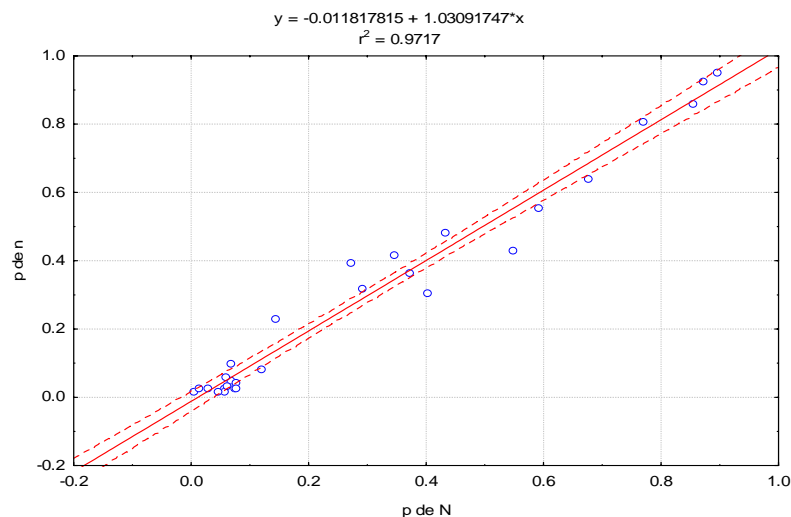


Fig. 2. Relación entre  $p$  de  $N$  (máximo valor de  $n$  con Eq. 1) y  $p$  de  $n$  (Eq. 2 y Fig. 1);  $z = 1.96$  y  $d = 0.05$ . Los puntos corresponden a 29 pares de datos de cuatro casos de insectos asociados a cafetales.

## Agradecimientos

Fundación Produce Chiapas financió este trabajo a través del proyecto "Bioecología y manejo de plagas del café en el Soconusco y Sierra de Chiapas".

### Literatura citada

- Daniel, W. W. 1981. Estadística con aplicaciones a las ciencias sociales y a la educación. McGraw-Hill. México, 504 p.
- Daniel, W. W. 2002. Bioestadística, base para el análisis de las ciencias de la salud. Limusa Wiley. México, 755 p.
- Zar, J.H. Biostatistical analysis. Fourth ed. Prentice Hall. EUA, 123 p.