Cuadernos de la S.E.C.F., N.º 1, octubre 1995, pp. 155-160 «Reunión de Valsaín. 17-20 octubre 1994»

EL MUESTREO 3P EN EL INVENTARIO FORESTAL

Guillermo Riesco Muñoz* & Juan Gabriel Álvarez González**

* Escuela Politécnica Superior. Universidad de Santiago de Compostela. 27002 LUGO
** Centro de Investigaciones Forestales de Lourizán. Apartado de Correos 127. 36080 PONTEVEDRA

1. INTRODUCCIÓN

El muestreo con probabilidad proporcional a la predicción (3P) se caracteriza porque cada árbol de la población tiene la posibilidad de ser elegido para pertenecer a la muestra con una probabilidad proporcional a alguna medida o valor arbitrario asignado al pie por métodos sencillos (por ejemplo, diámetro normal). Dicha medida se supone relacionada estadísticamente con la variable del árbol que interesa (por ejemplo, volumen maderable).

El muestreo 3P se aplica en aquellas poblaciones en que cada árbol ha de ser visitado una vez solamente. De hecho, se ha ideado para la valoración y venta de madera en pie, donde cada árbol que va a venderse debe ser marcado para la corta (GROSENBAUGH, 1965).

Al tiempo que se marca el árbol se le asigna el valor de una determinada variable (diámetro normal, altura total, volumen estimado ocularmente, etc.). Se compara ese valor con un número que se extrae en ese momento de una lista de números aleatorios. Si dicho número aleatorio es menor o igual a la variable medida el pie pasa la selección y entra a formar parte de la muestra, determinándose sobre él la variable de interés (normalmente el volumen real).

Se presenta el muestreo 3P como un método aplicable no sólo a la cubicación de madera en señalamientos sino también como segunda etapa en un muestreo bietápico (selección de pies en el interior de parcelas de muestreo). El ensayo se ha realizado con parcelas de 50 x 25 metros elegidas al azar sobre masas de pino radiata de 25 años de edad.

El número aleatorio usado en el criterio de selección de unidades muestrales se toma de una lista de números equiprobables comprendidos entre 1 y L. L debe superar o aproximarse al valor máximo de la variable medida sobre la población y se determina antes del muestreo. Existen programas de ordenador para obtener listas de números aleatorios entre unos límites dados (GROSENBAUGH, 1965; PARDO & DÍAZ, 1985). Para el presente estudio se ha diseñado la aplicación informática específica «NALEA», programada en CLIPPER 5.2c.

Los números aleatorios se obtienen con reemplazamiento porque todos ellos deben ser equiprobables, porque en general la frecuencia media absoluta con que aparece cada número debe ser superior a uno y porque la codificación del programa se simplifica.

2. ESTIMACIÓN DE EXISTENCIAS

Llamando x_i a la variable medida directamente en cada uno de los pies de la población y llamando L al mayor de los números aleatorios, la probabilidad de que un árbol i sea elegido para la muestra es $\pi_i = x_i/L$. Es decir, la probabilidad es proporcional a la magnitud de la variable medida.

Cuando los árboles más grandes son los que introducen más variabilidad en la población, se puede tomar L inferior al máximo valor de x_i. Así, todos los individuos cuyo valor de x_i sea igual o superior a L son seleccionados para la muestra con certeza y son considerados como un estrato aparte. El tamaño de muestra (n) aumenta y la precisión también ya que los pies seleccionados con certeza no incrementan el valor de la varianza.

Por ser un tipo de muestreo desarrollado para obtener los volúmenes totales de lotes de madera, los estimadores utilizados en muestreo 3P son los del volumen total de la población y se denominan estimador desajustado y estimador ajustado.

El estimador desajustado (Y_{dj})responde a la expresión:

$$Y_{dj} = \Sigma_{i=1...n}(y_i/\pi_i) = L * \Sigma_{i=1...n}(y_i/x_i)$$

donde x_i es la variable medida en todos los pies, y_i es la variable medida en la forma más precisa posible en los árboles seleccionados (normalmente el volumen) y n es el tamaño de la muestra.

Y_{dj} es un estimador muestral insesgado de $Y = \Sigma_{i=1..N} y_i$, donde N es el número de individuos de toda la población. La varianza del estimador Y_{dj} es:

$$V(Y_{dj}) = \Sigma_{i=1..N} y^2_i L/x_i - \Sigma_{i=1..N} y_i^2$$

que se puede estimar sin sesgo mediante la expresión:

$$v(Y_{dj}) = \Sigma_{i=1..n} y_i^2 * (L/x_i)^2 - \Sigma i=_{1..n} y_i^2 * L/x_i$$

La estimación muestral de la varianza es poco precisa debido al tamaño aleatorio y desconocido a priori de la muestra.

El estimador ajustado de Y se denomina Y_{aj} y se usa con más frecuencia que el desajustado en el muestreo 3P. Su expresión es la siguiente:

$$Y_{aj} = (\Sigma_{i=1..n} y_i/x_i) * \Sigma_{i=1..N} x_i/n$$

No hay expresiones rigurosas utilizables de la varianza del estimador ajustado. Para conocerla habrá que muestrear repetidamente una misma población y simular muestras por ordenador (GROSENBAUGH, 1965; PARDO & DÍAZ, 1985; SCHREUDER & *al.*, 1971). No obstante, la expresión más utilizada de la varianza poblacional es la propuesta por SCHREUDER (1971):

$$V(Y_{aj}) = (\Sigma_{i=1..N} X\, y_i/x_i - Y^2)/n_e$$

Un estimador insesgado de la misma es:

$$v(Y_{aj}) = \Sigma_{i=1..n}(X\, y_i/x_i - Y_{aj})^2/n(n-1)$$

Para las igualdades anteriores $X = \Sigma_{i=1..N} x_i$ y n_e es la esperanza matemática del tamaño de muestra n.

Al contrario que el estimador desajustado, el estimador ajustado utiliza toda la información obtenida en el muestreo (valores de x_i de toda la población y valores de y_i de la muestra). Además Y_{aj} es independiente de la arbitrariedad de L. Por ello la varianza $V(Y_{aj})$ es menor que $V(Y_{dj})$.

Y_{aj} presenta el inconveniente de ser un estimador sesgado de sesgo difícil de conocer y de tener una varianza desconocida al incluir en la expresión de la misma el tamaño de muestra, que es una variable aleatoria cuyo valor sólo se conoce al final del muestreo. Así, el error y coste de muestreo calculados a partir del tamaño de muestra esperado n_e pueden variar dependiendo del tamaño de muestra n realmente obtenido.

3. TIPOS DE VARIABLES Xi E Yi

Al final del muestreo se tiene el total poblacional de la variable relacionada x_i y el total muestral de la variable de interés y_i. Con estos valores se pueden calcular los estimadores y errores necesarios.

Los estimadores que se manejan en el muestreo 3P son los del total poblacional de alguna variable. De ahí que, para la estimación de las existencias de la masa se haya empleado como variable y_i el volumen unitario con corteza (Vcc) dado por la tarifa de dos entradas recogida en el II Inventario Forestal Nacional (ICONA, 1993) para pino radiata bien conformado en la provincia de Lugo:

$$Vcc = 33{,}43 + 0{,}0002719 * D^2 * H$$

donde Vcc es el volumen unitario con corteza en decímetros cúbicos, D es el diámetro normal con corteza en milímetros y H es la altura total en metros.

Un problema mayor es la elección de la variable x_i que se va a medir en cada uno de los pies. Esta debe tener fuerte correlación lineal con y_i porque así aumenta la precisión del estimador. No es necesario que dicha correlación sea conocida pero sí es recomendable que exista. Asimismo, la recta de regresión y_i frente a x_i debe estar poco separada del origen. En la fórmula propuesta para Vcc se trataría de que H tuviera poca variabilidad y que el término independiente fuera pequeño.

El valor de x_i ha de ser de obtención sencilla y barata por el hecho de que debe determinarse en todos los individuos de la masa. Se ha elegido como variable x_i el diámetro normal con corteza al cuadrado, siguiendo a PARDO & DÍAZ (1985), debido a su correlación con Vcc.

4. DESARROLLO DEL PLAN DE MUESTREO

4.1. Determinación de L

L es la parte entera de X/n_e, donde debe hacerse una estimación previa de X mediante un muestreo piloto. El tamaño esperado de muestra n_e se obtiene con la expresión del muestreo aleatorio simple:

$$n_e = (t * c.v./d)^2$$

aplicable si la fracción de muestreo es despreciable frente a la unidad (inferior al 5%). En dicha expresión d es el error relativo de r respecto de la media de cocientes poblacional, t es el valor de la variable normal N(0,1) para el nivel de confianza que se pretende (usualmente el 1, 5 ó 10%) y c.v. es el coeficiente de variación de ri, que responde a la expresión:

$$c.v. = S/r$$

donde

$$r = \Sigma i =_{1...n} y_i/(x_i*n)$$

$$S^2 = \Sigma_{i=1...n}(r_i-r)^2/(n-1)$$

$$r_i = y_i/x_i$$

Como no se conoce ningún valor de y_i hasta despues de realizado el muestreo, en la práctica el coeficiente de variación se debe predecir. Su valor oscila entre el 15 y el 30%.

4.2. Programa generador de números aleatorios

Un conjunto de números es aleatorio cuando todos ellos tienen la misma probabilidad de ocurrencia.

Se ha diseñado en CLIPPER 5.2c el programa «NALEA», que permite conseguir una lista tan extensa como se desee de números enteros aleatorios comprendidos entre dos valores cualesquiera. El ordenador dispone de una función generadora de números de cuatro cifras, entre el 0000 y el 9999, llamados «números pseudoaleatorios» dado que son generados por algoritmos matemáticos iterativos y no de forma estrictamente arbitraria (AZORIN, 1972). El programa «NALEA» transforma, mediante una expresión matemática lineal, cada serie de cuatro cifras (Z) en un número aleatorio entero (z) comprendido dentro de un intervalo cuyos valores extremos son N_0 y N_1, dos números enteros cualesquiera donde N_0 es menor o igual que N_1. La expresión matemática propuesta es la siguiente:

$$z = E[Z * (N_1 - N_0 + 1)/10000 + N_0]$$

El número z es entero dado que en el muestreo 3P se va a comparar con lecturas de forcípula, cuya precisión es de 0'5 centímetros. Si x_i es el diámetro normal al cuadrado $N_0 = 1$ y N_1 es $E(\sqrt{L})$ (parte entera de $\sqrt{L}$). Esto permite comparar directamente en el monte diámetros normales en cm. con números aleatorios en cm.

El procedimiento de conversión de Z en z se repite para cada número hasta que la lista tiene la cantidad de números prefijada, que será superior o igual al total de árboles de todas las parcelas de un mismo estrato homogéneo de la masa a inventariar.

El ordenador presenta la ventaja de generar números aleatorios con gran rapidez en el intervalo de valores que se precise y en la cantidad necesaria, pero presenta asimismo el inconveniente de que es necesario contrastar estadísticamente la aleatoriedad de las secuencias de números aleatorios que se obtengan. Existen varias clases de pruebas especialmente diseñadas para este fin, que garantizan en lo posible la ausencia de tendencias o sesgos. El programa efectúa la prueba siguiente.

Dada una lista de k números enteros aleatorios con reemplazamiento $z_1...z_k$, no menores que N_0 y no mayores que N_1, la frecuencia absoluta esperada de un número entero N_j del intervalo $[N_0, N_1]$, donde j varía de 1 a $E(\sqrt{L})$, será:

$$k/(N_1 - N_0 + 1)$$

Este valor se compara con la frecuencia absoluta $f(N_j)$ realmente obtenida para cada número N_j. La lista es aleatoria al nivel de confianza establecido si la diferencia entre ambas frecuencias, la real y la esperada, es aproximadamente nula. La diferencia puede estudiarse utilizando la χ^2 de Pearson, que se calcula como suma de cocientes de los cuadrados de las desviaciones entre frecuencias absolutas y esperadas, divididos por las frecuencias esperadas. En este caso:

$$\chi^2 = \Sigma_{j=1..E(\sqrt{L})}[f(N_j) - k/(N_1 - N_0+1)]^2 * *(N_1 - N_0+1)/k = (N_1 - N_0+1)/k * *\Sigma_{j=1..E(\sqrt{L})} f^2(N_j) - k$$

Se comprueba si el valor obtenido de la expresión anterior es significativamente grande (mayor que el dado por la tabla de la χ^2 con $N_1 - N_0$ grados de libertad) en cuyo caso se rechaza la lista como no aleatoria. La comprobación se efectúa para un determinado nivel de confianza.

La distribución $\sqrt{(2\chi^2)}$ se aproxima sensiblemente a la normal N $\{\sqrt{[2(N_1 - N_0) - 1]}, 1\}$ cuando el número de grados de libertad $N_1 - N_0$ es mayor que 30 (RÍOS, 1985). Es decir, χ^2 toma un valor muy próximo a :

$$0{,}5\{a + \sqrt{[2(N_1 - N_0) - 1]}\}^2$$

donde a es -1,282, -1,645, -2,326 ó -2,575 para un nivel de significación del 10, 5, 1 ó 0,5% respectivamente. Cuando $N_1 - N_0$ es menor o igual que 30 se emplean las tablas de la χ^2 almacenadas en el programa.

Según STEEL & TORRIE (1981), citando a Cochran, la frecuencia absoluta esperada para que la prueba sea aplicable debe ser superior a 5, es decir, $k/(N_1 - N_0+1)$ mayor que 5.

4.3. Fase de campo

El coste de los trabajos de campo es el más importante de un plan de muestreo. En el caso del muestreo 3P se tiene el coste añadido de generar una lista de números aleatorios y de alguien que la lleve al monte. Se debe idear algún sistema para la extracción de los números en el campo. GROSENBAUGH(1965) propuso un procedimiento basado en cintas de papel enrolladas con los números aleatorios impresos.

Se debe cuidar que el encargado de medir o hacer juicios de las variables no guarde ninguna relación con el encargado de proporcionar los números aleatorios. Éste cantará los números siempre después de efectuada la medida o estimación, a fin de reducir el sesgo.

5. DISCUSIÓN

El muestreo 3P es un método aplicable no sólo a la cubicación de madera en señalamientos sino también como segunda etapa en un muestreo bietápico. En cada una de las parcelas elegidas según un plan clásico de muestreo sistemático se mide el diámetro normal de todos los pies y en algunos de ellos, elegidos habitualmente de forma subjetiva, se mide la altura para proceder a la cubicación. Se ha sustituido dicho muestreo subjetivo de la segunda etapa por un muestreo estadístico 3P donde todos los árboles de todas las parcelas constituyen la población y sólo sobre los seleccionados se mide la altura total.

El ensayo se ha realizado en el área recreativa de Lagos de Teixeiro, en las proximidades de la ciudad de Lugo, con parcelas de 25 x 50 metros elegidas al azar, ubicadas y replanteadas sobre masas de pino radiata de 25 años de edad, con densidad media de 1500 pies por hectárea y 23 metros de altura dominante.

En el muestreo 3P polietápico, elaborado por Grosenbaugh en 1971, se toma en una primera fase una muestra con probabilidad proporcional al diámetro normal al cuadrado (D^2). En ella se selecciona un conjunto de pies con probabilidad proporcional a la altura maderable (Hmad) estimada. Sobre este conjunto, seleccionado respecto de toda la población con probabilidad proporcional a D^2*Hmad, se efectúa la medida rigurosa del volumen mediante aparatos. Según datos de VAN HOOSER (1972), los resultados así obtenidos tienen un error muy similar al que se obtendría por muestreo por regresión con muestras mucho mayores.

WIANT (1977) propone una variante. Cuando sobre las parcelas de la primera fase se han medido o estimado las alturas de todos los árboles se puede seleccionar una submuestra de árboles con probabilidad proporcional a su altura o bien se seleccionan unidades de muestreo puntual con probabilidad proporcional a la suma de alturas de los pies en ellas incluidos. WIANT comparó ambos sistemas con un muestreo aleatorio de igual tamaño muestral. El método de suma de alturas es más preciso y da los mayores rendimientos en la fase de campo.

En muestras grandes el 3P es más eficiente que el muestreo aleatorio pero el muestreo sistemático es superior en cuanto a la varianza del estimador ajustado ya que el tamaño aleatorio de la muestra introduce variabilidad adicional. El coste esperado es similar en muestreo 3P y sistemático (BONNOR, 1972)

6. CONCLUSIONES

La precisión y exactitud aumentan en masas regulares y semirregulares (variación de x_i e y_i pequeña) y si la diferencia entre el tamaño real de la muestra y el esperado ($n - n_e$) es mínima.

Diversos autores (EK, 1971; SCHREUDER *& al.*, 1971) han indicado que la precisión del muestreo 3P es baja debido a que no hace uso de la relación estadística existente entre x_i e y_i y que sería conveniente hacer uso de esta información. Si hay fuerte correlación entre la variable de interés y la efectivamente medida el estimador es débilmente afectado por el tamaño muestral. Como dicho tamaño es una variable aleatoria, si la correlación no es grande, habrá una variabilidad adicional que afectará al estimador. Es decir, interesa que la varianza de $r_i = y_i/x_i$ sea baja. El muestreo 3P subestima la varianza real si las muestras son pequeñas y lo contrario si son grandes.

El otro gran inconveniente del muestreo 3P es el tamaño aleatorio de la muestra, que hace que coste y precisión sean distintos de los planeados, lo cual limita fuertemente la aplicabilidad del método.

En una fase más avanzada del estudio se pretende obtener estimadores insesgados de los parámetros poblacionales tomando dos o más muestras independientes con reemplazamiento (GROSENBAUGH, 1965). Esto se consigue comparando la variable medida x_i con dos o más números aleatorios y aplicando el criterio de selección 3P. El estimador ajustado será la media ponderada según los tamaños obtenidos de las muestras y el estimador desajustado será la media aritmética de los obtenidos en cada muestra.

REFERENCIAS BIBLIOGRÁFICAS

AZORÍN, F.; 1972. *Curso de muestreo y aplicaciones*. Ed. Aguilar. Madrid.

BONNOR, G. M.; 1972. A test of 3P sampling in Forest Inventories. *Forest Science*, 18 (3): 98-202.

DÍAZ, M.; 1980. *Técnicas de muestreo forestal.* Escuela Técnica Superior de Ingenieros de Montes. Madrid.

EK, A. R.; 1971. A comparison of some estimators in forest sampling. *Forest Science*, 17 (1): 2-15.

GROSENBAUGH, L. R.; 1965. *Three-pee sampling theory and program THRP for computer generation of selection criteria. USDA*. For. Ser. Res. Pap PSW-21.

ICONA: 1993. *Segundo Inventario Forestal Nacional 1986-1995. Galicia. Lugo.* Ministerio de Agricultura, Pesca y Alimentación. Madrid.

PARDO, M. & F. DÍAZ; 1985. *Muestreo 3P: estudio de la posibilidad de esta técnica como submuestreo en parcelas permanentes de un I.F.C.* Documento sin publicar.

PITA, P. A.; 1973. *El Inventario en la Ordenación de montes.* INIA. Madrid.

RÍOS S.; 1985. *Métodos estadísticos. 2ª ed.* Ediciones del Castillo, S. A. Madrid.

SCHREUDER, H. T., J. SEDRANSK & K.D. WARE; 1968. 3P sampling and some alternatives, I. *Forest Science,* 14 (4): 429-454.

SCHREUDER, H. T., J. SEDRANSK & K.D. WARE; 1971. 3P sampling and some alternatives, II. *Forest Science*, 17 (1): 103-118.

STEEL, R. G. & J.H. TORRIE; 1981. *Principles and procedures of statistics: a biometrical approach.* Mc Graw-Hill. New York.

VAN HOOSER, D. D.; 1972. E*valuation of two-stage 3P sampling for forest surveys.* USDA For. Ser. Res. Pap SO-77.

WIANT, H. V.; 1977. *Comparison of point-3p sampling designs*. Resource Inventory Notes. USDI D-340.