Diseños muestrales en Inventarios Forestales

Introducción

Se entiende aquí por Diseño Muestral, el procedimiento para seleccionar individuos (Unidades Muestrales) de una población (Unidad de Inventario) y definir en base a ellos, indicadores de su estado (parámetros de Rodal o Variables de Estado).

No se pretende agotar aquí el tema sobre diseños muestrales empleados en los inventarios forestales. Su objetivo principal es describir los diseños más comunes usados en los inventarios que se ejecutan con el fin de tomar decisiones, en especial respecto al manejo del bosque y aprovechamiento.

Quedan, por lo tanto, fuera del alcance de este documento, los diseños que se refieren a inventarios cuyo fin es definir o evaluar políticas forestales, regionales o nacionales, los que por su complejidad requieren de una base estadística y tecnológica de alta especialización.

Atendiendo a la definición antes dada, es necesario referirse a los siguientes elementos:

Unidades de Inventario

El primer paso al diseñar un inventario es la división del bosque en Unidades de Inventario (UI), es decir, en las superficies boscosas básicas en que se localizará un conjunto independiente de unidades muestrales, con las que se describirá la Unidad mediante una serie de Variables de Estado estimadas.

Este tema se trata en un documento diferente (**Elementos de un Inventario Forestal**)

Unidades Muestrales

En la descripción del bosque mediante muestreo, los individuos poblacionales (o unidades muestrales) no son arboles individuales, sino conjuntos de árboles llamados Unidades Muestrales (UM) o vulgarmente parcelas. Este tema también es tratado en un documento diferente (**Unidades Muestrales**)

Parámetros o Variables de Estado

Los datos de los arboles seleccionados en las UM que representan estadísticamente a la UI, son procesados de manera que se obtiene, indicadores del estado del bosque, llamados ordinariamente Parámetros de Rodal o Variables de Estado (VE).

Estos elementos son tratados en un documento aparte (**Variables de Estado).**

Estimadores

Cada diseño muestral, dependiendo de la naturaleza de las UM y de la forma en que estas se distribuyen, tiene sus propios estimadores.

Un estimador describe una VE, que puede ser un Total Poblacional, un Total por Hectárea, o un Valor Promedio de algún atributo de los árboles que integran el bosque o UI.

Los estimadores poseen una serie de propiedades estadísticas que son valorizadas en forma diferente por cada diseño muestral:

Sea $\hat{X}$ estimador de una VE, cuyo valor real es X. Si el proceso de estimación muestral se repite, se obtienen nuevos y diferentes valores de $\hat{X}$ que constituyen una población que puede llegar a contener infinitos estimadores del mismo valor X.

Se dice que el estimador $\hat{X}$ es insesgado cuando su Esperanza Matemática, es decir, el promedio de todos los valores $\hat{X}$ que constituyen la población de estimadores, es igual al valor real X:

$$E\left[\hat{X}\right] = X$$

El error de un estimador $\hat{X}$ respecto al valor real X se expresa mediante el Error Cuadrático Medio:

$$ECM(\hat{X}) = E\left[\hat{X} - X\right]^2$$

La variabilidad de un estimador se expresa, en cambio, por la Varianza:

$$V(\hat{X}) = E\left[\hat{X} - E\left[\hat{X}\right]\right]^2$$

Si un estimador es insesgado, es decir, si X es igual a $E\left[\hat{X}\right]$, su Error Cuadrático Medio es igual a su varianza.

En base a una muestra aleatoria de Unidades Muestrales, es posible **estimar** la varianza de los estimadores, en cambio el ECM no es posible estimarlo, pues el valor real X será siempre desconocido.

Si los estimadores son insesgados, al menos para los efectos prácticos, esta varianza estimada se puede emplear como expresión del error del estimador que se simboliza como $S_{\hat{X}}$ y que se llama vulgarmente **Error de Estimación**:

$$S_{\hat{X}}^2 = \hat{V}\left(\hat{X}\right)$$

Error Muestral

Los estimadores basados en muestras aleatorias tienden a ser insesgados y a distribuirse conforme a la **Distribución Normal**. La condición de insesgamiento y de aproximación a la distribución normal permiten usar como expresión del error del estimador, el llamado **Error Muestral.** El Error Muestral ($E_{\hat{X}}$) es la máxima diferencia esperada entre el valor real de una VE y su valor estimado en base a una muestra, para una probabilidad o nivel de confianza dado.

Intervalo de Confianza

La Expresión $\hat{X} \pm E_{\hat{X}}$ se denomina el Intervalo de Confianza del estimador $\hat{X}$ y define el rango dentro del cual se encuentra el verdadero valor X con una probabilidad o nivel de confianza dado.

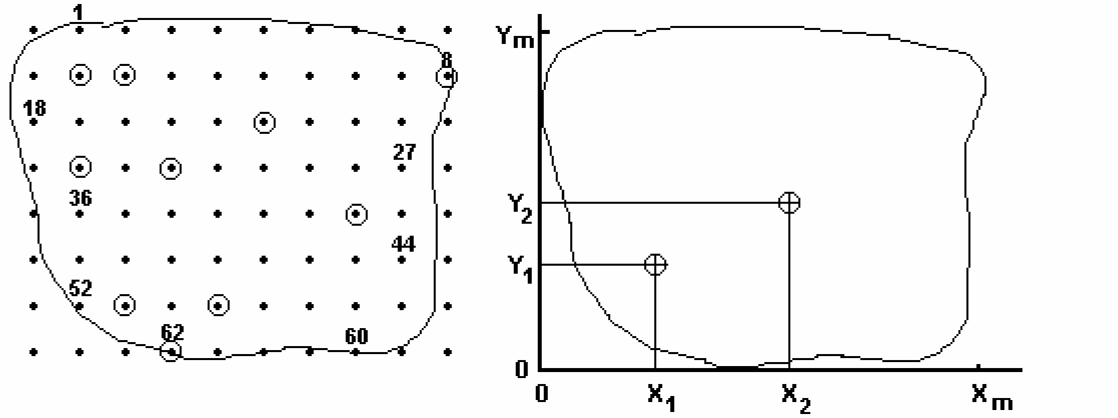
El fundamento estadístico de los estimadores y de sus propiedades están fuera del alcance de este documento. Para ello el lector puede referirse a textos como De Vries(1986), Raj(1968) o Cochran (1977)

Distribución de las unidades muestrales.

Distribución Aleatoria

Los procedimientos de estimación de parámetros que se verán más adelante, asumen que las u.m. se distribuyen **aleatoriamente** en el rodal. Sin embargo, se ha podido comprobar en numerosas ocasiones que al emplear otras formas de distribución, puede lograrse una mayor eficiencia, y el hecho de no existir aleatoriedad no impide adoptar los estimadores y sus correspondientes errores asociados a un muestreo aleatorio.

La figura 1 muestra dos procedimientos para distribuir unidades muestrales al azar.



a) Localización aleatoria de n=10 parcelas seleccionando en una red sistematica de N=62 puntos:

8	15	16	23	32	34	42	53	55	62

b) Localizacion aleatoria de parcelas mediante coordenadas.

Figura 1

La figura 1a presenta la forma más común de localización, consistente en una aleatoriedad restringida. Se genera una red sistemática de puntos, con una distancia entre puntos en la red igual a la distancia mínima a la que deben ubicarse los puntos de muestreo en el terreno**.** Luego se selecciona una muestra aleatoria de puntos de la red que se materializan en terreno como centros de parcelas.

La figura 1b muestra el procedimiento de localización aleatoria estricta. Se enmarca la Unidad de inventario en un sistema de coordenadas rectangulares y se selecciona un par de coordenadas aleatorias (x,y), Al emplear este procedimiento, los centros de parcelas pueden localizarse muy cercanos, incluso coincidentes , de manera que dos Unidades Muestrales pueden contener los mismos árboles. La localización restringida (1a) evita un evento de este tipo.

La localización aleatoria de Unidades Muestrales es probablemente la forma más ineficiente de distribución muestral, pero es la única que puede garantizar la representatividad estadística y con ello la condición de insesgamiento de los estimadores.

Distribución Sistemática

La distribución llamada sistemática consiste en diseñar una red de parcelas distribuidas conforme a un arreglo regular, que puede ser cuadrado, rectangular, triangular o cualquiera. Luego se elige un punto (partida) aleatorio en la UI, el que se hace coincidir con cualquier punto de la red. Esta forma de distribución (figura 2) genera estimaciones insesgadas debido a la aleatoriedad de la partida. Sin embargo, en estricto sentido estadístico, toda la red de unidades constituye un sola UM, por lo que es imposible calcular el error de los estimadores.

Se han propuesto varias formas de aproximarse al error de estimación, las que responden a diversos supuestos. El supuesto mas común es el de población en orden aleatorio. Si este supuesto fuera cierto, los estimadores de error correspondientes al Muestreo Aleatorio Simple (MAS) son plenamente válidos aun cuando la muestra se distribuya sistemáticamente. Para las intensidades usuales de los inventarios de rodal, un sesgamiento grande de los estimadores de error es muy improbable. Por el contrario, se ha podido comprobar que generalmente el error de estimación determinado bajo el supuesto de muestreo aleatorio sobrestima el verdadero error muestral en una pequeña magnitud, lo que viene a constituir un factor de seguridad dado que se construyen intervalos de confianza un poco mas amplios de lo que en realidad son.

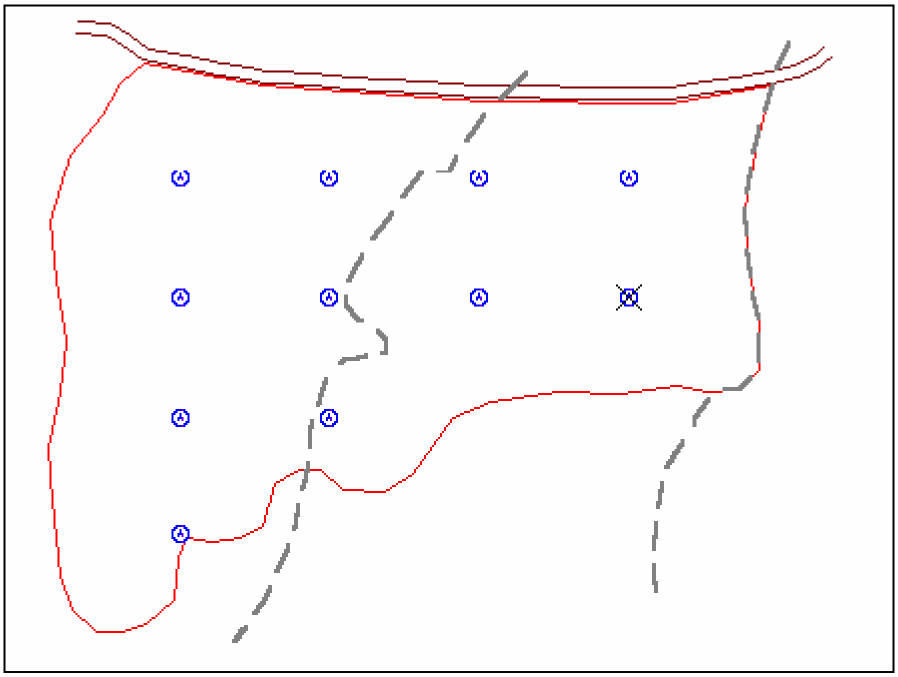


Figura 2. Distribución muestral sistemática con partida aleatoria

En Prodan el al.(1997) pueden verse otras formas de realizar estimaciones de parámetros y de sus errores, para distribución de UM. en redes sistemáticas.

Distribución de las UM en transectos

Transectos lineales

El procedimiento consiste en fijar una línea que atraviese el rodal de un extremo a otro. La línea de longitud L se divide en n segmentos de longitud l metros. Luego se elige una partida al azar l_0, entre 1 y l, donde se ubica la primera parcelas y las restantes a las distancias l_0+l, l_0+2l, etc. (5a). Este procedimiento puede dar resultados satisfactorios cuando la mayor variabilidad en el rodal se da en su parte central.

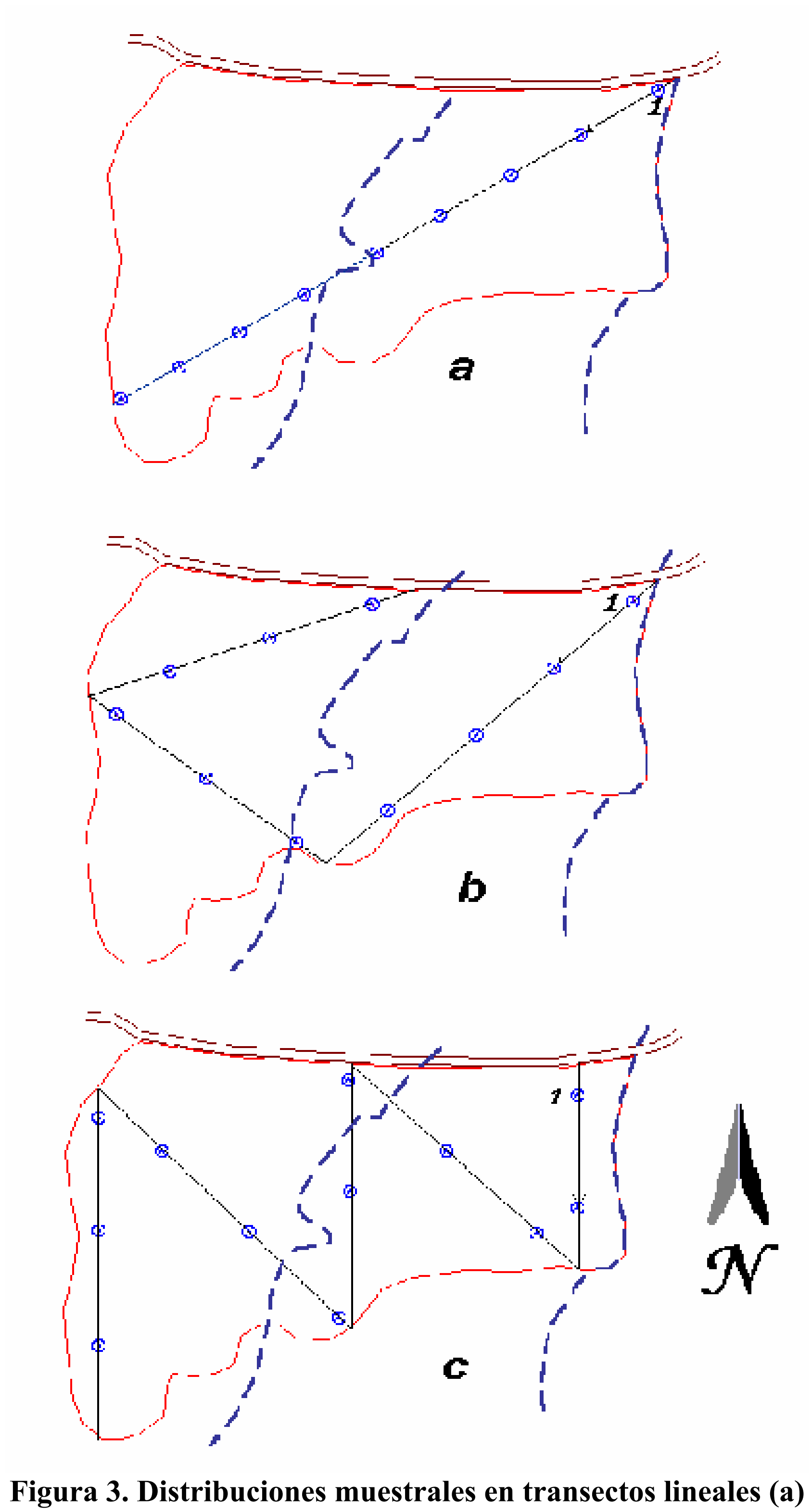


Figura 3. Distribuciones muestrales en transectos lineales (a) Recorrido estructurado (b) y Ricochet (c)

Transecto de Ricochet

Es una modificación del transecto lineal. El transecto de Ricochet (figura 3b) parte de un punto localizado al azar, en una dirección cardinal fijada según la forma del rodal. El transecto avanza hasta tocar un borde, donde rebota en un ángulo incrementado en 45 grados. Deben evitarse los cruces de segmentos del transecto. La distancia entre parcelas se establece en la siguiente forma (Lund y Thomas, 1989):

$$l = 107.5\sqrt{(A / n)} \tag{10}$$

Donde A es la superficie del rodal en hectáreas y n el numero de parcelas que se distribuyen. Las parcelas se ubican a lo largo del transecto en la misma forma que para el transecto lineal.

Transecto en Recorrido Estructurado

Maclaren y Goulding (1993) proponen una forma de distribución de parcelas en rodales pequeños (hasta aprox. 40 ha) que llaman el recorrido estructurado(figura 3c). Consiste en fijar un recorrido que, con solo uno o dos quiebres de dirección, cubra al rodal uniformemente y comience y concluya en puntos logísticamente favorables. Las parcelas se ubican con partida aleatoria y equidistantes a lo largo del recorrido como en el transecto de Ricochet. Al establecer el recorrido, debe minimizarse el riesgo de coincidencia entre éste y posibles estructuras naturales o construidas, como caminos, quebradas, etc. La posición de la primera parcela puede ser aleatoria con lo que todo el conjunto de parcelas logra una mejor representatividad.

Estimadores para unidades muestrales de igual tamaño

Cuando las Unidades Muestrales (UM) son todas **del mismo tamaño**, se emplean los estimadores de **Muestreo Aleatorio Simple (MAS).**

Se considera que las UM son del mismo tamaño cuando a cada arbol de una UI se asigna la misma probabilidad de ser seleccionado en todas las UM que se localizan en ella, aun cuando esta probabilidad puede variar de arbol en arbol. Si se emplean, por ejemplo, parcelas circulares, todas deben tener la misma superficie. Si se emplean, en cambio, parcelas de radio variable (Bitterlich o MPH) todas deben tener el mismo factor **K** que multiplica al diámetro del árbol para definir su area de selección, etc. Los estimadores MAS se adoptan también cuando se emplean UM en conglomerado, siempre que estos incluyan un número constante de parcelas y estas sean todas del mismo tamaño.

En estrictos términos estadísticos, los estimadores MAS asumen que las UM se han localizado aleatoriamente en la población. No obstante, ellos se usan también cuando se emplean otros tipos de distribución muestral, en cuyo caso no existirá certeza de insesgamiento respecto a las estimaciones y a sus errores.

Los estimadores MAS que se verán aquí no son los mismos que se pueden encontrar en los textos que en general describen a estos estimadores. Cuando se definen los estimadores MAS se asume que el marco poblacional es finito, es decir, que el número total de individuos (o Unidades Muestrales) que integran la población es conocido y finito, = N.

Estimadores MAS tradicionales

Sobre una población que contiene **N** individuos se seleccionan **n** de ellos en forma tal, que todas las posibles muestras de tamaño n tienen la misma probabilidad de ser elegidas. En la práctica se logra el mismo efecto si los individuos que integran la muestra se selecciona uno a uno dando a todos la misma probabilidad de ser elegidos en cada oportunidad. Esta selección puede efectuarse en dos formas: un individuo seleccionado tiene oportunidad de serlo nuevamente (Con Reemplazo) o no (Sin Reemplazo).
A cada individuo (i) seleccionado se le mide un atributo cualquiera (x_i)

Promedio Poblacional

El valor promedio poblacional de un atributo cualquiera **x** de los individuos de una población es estimado mediante el promedio aritmético del mismo atributo de los individuos que integran la muestra:

$$\bar{X} = \sum^{N} x / N \hat{=} \bar{x} = \sum^{n} x / n = \hat{\bar{X}}$$

Esta expresión se interpreta del siguiente modo: La media poblacional ($\bar{X} = \sum^{N} x / N$) es estimada **insesgadamente** ($\hat{=}$)mediante la media muestral $\bar{x} = \sum^{n} x / n$. A este estimador se le llama también $\hat{\bar{X}}$ o estimador de la media poblacional

Total poblacional

El valor total poblacional del atributo **x** es estimado mediante una muestra aleatoria de tamaño n del siguiente modo:

$$X = N \cdot \bar{X} \hat{=} N \cdot \bar{x} = \hat{X}$$

Varianza de las Medias Muestrales

(o de los estimadores de la media poblacional $\bar{X}$

En términos puramente teóricos, la varianza de las medias muestrales puede determinarse en la siguiente forma, que supone conocer el valor **x** de todos los **N** individuos que integran la población:

$$V(\hat{\bar{X}}) = \frac{S_x^2}{n}\left(1 - \frac{n}{N}\right) \quad (Sin \;\; reemplazo)$$

$$= \frac{\sigma_x^2}{n} \doteq \frac{S_x^2}{n} \quad (Con \;\; reemplazo)$$

Donde S_x^2 es la **Varianza poblacional corregida:**

$$S_x^2 = \sum^{N} (x - \bar{X})^2 / (N - 1)$$

La Varianza Poblacional describe la variabilidad entre individuos de la población. Esta también se acostumbra a representar mediante el Coeficiente de Variación:

CV(x) = ($S_x / \bar{X}$)*100

Error de Estimación del Promedio

La Varianza de los estimadores de la media poblacional ($V(\hat{\bar{X}})$) nunca es conocida exactamente pues se desconoce la varianza poblacional S_x^2. Sin embargo, ella puede estimarse, también insesgadamente, mediante la varianza muestral s_x^2:

$$\hat{V}(\hat{\bar{X}}) = S^2_{\hat{\bar{x}}} = \frac{s_x^2}{n}(1-\frac{n}{N}) \quad (Sin \;\; reemplazo)$$

$$= \frac{s_x^2}{n} \quad (Con \;\; reemplazo)$$

Puesto que la media muestral estima insesgadamente a la media poblacional, La **Varianza estimada** de este estimador es al mismo tiempo un estimador insesgado del error cuadrático medio del estimador. Por ello se le llama simplemente **Error de Estimación de la media:**

$$S^2_{\hat{\bar{x}}} = \frac{s_x^2}{n}(1-\frac{n}{N}) = \frac{\sum^n (x-\bar{x})^2}{(n-1)n}\left(1-\frac{n}{N}\right) = \frac{\sum^n x^2 - \left(\sum^n x\right)^2 / n}{(n-1)n}\left(1-\frac{n}{N}\right)$$

El término $\left(1-\frac{n}{N}\right)$ se llama factor de corrección para muestreo sin reemplazo. En la práctica solo se emplea cuando la fracción muestral (n/N) es relevante, es decir cuando (n/N) > 0.05

Error de Estimación del Total Poblacional

El error de estimación del total poblacional se determina en la siguiente forma:

$$S^2_{\hat{X}} = N^2 S^2_{\bar{x}} = N^2 \frac{s_x^2}{n}(1-\frac{n}{N})$$

Error Muestral del Promedio

Se llama Error Muestral ($E(\hat{\bar{X}})$)a la máxima diferencia esperada entre el valor estimado $\hat{\bar{X}}$ y el correspondiente valor poblacional real $\bar{X}$, para un nivel de confianza dado.
En muestras grandes (n > 30) su valor puede determinarse gracias a que en tales circunstancias, los estimadores $\hat{\bar{X}}$ tienden a distribuirse **normalmente .** El Error Muestral, entonces, se aproxima al siguiente valor, como se demuestra en cualquier texto de estadística muestral donde se describen los sustentos teóricos del MAS:

$$E(\hat{\bar{X}}) = z_{1-\alpha/2} S_{\hat{\bar{X}}}$$

Donde $z_{1-\alpha/2}$ es un múltiplo del Error de Estimación, tal, que existe solo una probabilidad α que la diferencia entre $\hat{\bar{X}}$ y $\bar{X}$ sea mayor que $E(\hat{\bar{X}})$.
Esta propiedad, sin embargo, no puede emplearse en la práctica, pues supone que la muestra es grande y que la Varianza de $\hat{\bar{X}}$ es conocida y no solamente su estimador $S^2_{\hat{\bar{X}}}$.
En la práctica esto significa que el valor de $z_{1-\alpha/2}$ debe ser sustituido por el valor $t^{n-1}_{1-\alpha/2}$, de manera que en la práctica la expresión del error muestral es:

$$E(\hat{\bar{X}}) = t_{1-\alpha/2}^{n-1} S_{\hat{\bar{X}}}$$

Los valores t varían dependiendo del tamaño de la muestra n y del **Nivel de Confianza 1-** α .Se encuentran tabulados en la mayoría de los textos de estadística. El siguiente es un pequeño extracto de una tabla de valores t, tomada de Rohlf y Sokal (1969):

Tabla 1 Valores críticos de la distribución t de Student

Grados de Libertad (n-1)	**Valor crítico de la distribución de t para diferentes valores de α**				
	α**=0.4**	**0.2**	**0.1**	**0.05**	**0.01**
1	1.376	3.078	6.314	12.706	63.657
5	0.920	1.476	2.015	2.571	4.032
10	0.879	1.372	1.812	2.228	3.169
15	0.866	1.341	1.753	2.131	2.947
20	0.860	1.325	1.725	2.086	2.845
30	0.854	1.310	1.697	2.042	2.750
60	0.848	1.296	1.671	2.000	2.660
120	0.845	1.289	1.658	1.980	2.617
∞	0.842	1.282	1.645	1.960	2.576

El error muestral puede expresarse, también en forma relativa o %:

$$E\% = \frac{E(\hat{\bar{X}})}{\hat{\bar{X}}} \cdot 100 = \frac{E(\hat{X})}{\hat{X}} \cdot 100$$

Límites Confidenciales

Se llaman Límites Confidenciales o Intervalo de Confianza al rango generado al sumar y restar al valor estimado $\hat{\bar{X}}$, el Error Muestral :

$$\text{LC}(\hat{\bar{X}}) = \hat{\bar{X}} \pm E(\hat{\bar{X}}) = \hat{\bar{X}} \pm t_{1-\alpha/2}^{n-1} S_{\hat{\bar{X}}}$$

Los Límites Confidenciales representan el intervalo dentro del cual se encuentra el verdadero valor poblacional, con una probabilidad dada (= 1 - α)

Existe una probabilidad α /2 de un valor real $\bar{X}$ que exceda el limite superior del intervalo y la misma probabilidad de que sea menor que el limite inferior del rango.

Los límites Confidenciales pueden definirse también en función del Error Muestral %:

$$LC(\hat{\bar{X}}) = \hat{\bar{X}} \cdot (1 \pm E\%/100)$$

Del mismo modo se pueden definir para el total poblacional:

$$LC(\hat{X}) = \hat{X} \cdot (1 \pm E\%/100)$$

Ejercicios

En la figura 4 se encuentre la **Población de Juego I** que representa un bosque de 40 hectáreas dividido en 400 parcelas cuadradas de 1000 m^2 cada una. El valor registrado en cada casillero representa el volumen sumado de todos los árboles medidos en cada parcela, en décimos de m^3.

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	
1	130	153	153	112	200	106	100	147	118	165	0	0	12	0	35	0	18	0	0	24	
2	124	106	136	130	165	141	194	212	136	88	100	0	12	65	88	0	100	30	12	47	
3	177	165	136	124	171	106	82	177	147	165	118	82	47	6	88	12	30	0	0	24	A
4	165	112	124	118	153	118	224	136	118	159	141	65	35	24	0	30	30	53	53	30	
5	100	82	118	153	147	130	130	112	88	118	147	153	88	53	71	0	0	94	47	30	
6	224	247	217	230	130	259	277	100	147	171	200	171	118	141	82	59	71	6	0	0	
7	253	200	135	271	277	271	230	206	242	177	141	200	135	153	106	153	124	71	30	6	
8	212	277	265	212	206	171	289	259	183	247	194	277	183	165	88	106	118	136	53	71	B
9	224	283	247	300	100	318	277	306	177	200	177	271	141	71	124	71	188	171	159	94	
10	100	141	265	277	306	165	253	265	271	159	236	188	300	165	147	241	118	159	82	124	
11	277	330	253	218	177	353	330	253	171	194	241	177	177	118	88	106	118	188	77	165	
12	224	212	159	224	141	183	283	188	147	183	206	183	130	88	59	130	141	112	106	94	
13	271	318	200	271	218	253	260	200	147	259	253	77	165	242	153	194	106	224	59	141	C
14	277	277	206	236	230	230	294	165	294	212	259	159	94	124	212	100	159	124	218	200	
15	130	218	65	171	165	194	171	206	312	94	153	118	171	71	136	147	88	100	153	124	
16	218	130	118	130	82	171	147	124	177	183	159	94	124	212	100	159	124	100	82	71	
17	106	147	153	118	159	153	153	130	112	177	88	12	41	18	24	88	53	41	0	18	
18	130	200	194	100	141	165	153	147	177	194	106	35	0	18	0	0	35	30	41	35	D
19	77	165	159	159	183	118	124	124	94	159	71	0	100	18	6	6	0	0	0	30	
20	188	183	177	130	94	153	47	188	112	118	18	18	0	0	0	12	0	30	59	12	
			I					II					III					IV			

Figura 4. Selección aleatoria de 30 UM sobre

Población de Juego I

La Tabla 2 presenta el resultado del muestreo, comparado con los valores poblacionales reales

Tabla 2

Parámetro	**Valor Poblacional**		**Estimador**	
	Simbolo	**Valor**	Símbolo	**Valor**
Promedio	$\overline{X}$	**13.6425**	$\hat{\overline{X}}$	**13.100**
Total	X	**5457**	$\hat{X}$	**5420**
Varianza	S_x^2	**66.765**	s_x^2	**53.14**
Coeficiente de Variación	CV(x)%	**59.82%**	cv(x)%	**55.65**
Varianza de las medias Muestrales	**V($\hat{\overline{X}}$)**	**2.059**	$S^2_{\hat{\overline{X}}}$	**1.64**
	$\left(\sqrt{V(\hat{\overline{X}})}\right)\%$	**10.51%**	S%	**9.77**
Error Muestral	$E(\hat{\overline{X}})$	**2.92 (21.42%)**		**2.61 (19.94%)**

Estimadores MAS Alternativos

Los estimadores MAS tradicionales solo excepcionalmente pueden ser empleados en Inventarios Forestales. Las Unidades Muestrales (UM) empleadas normalmente en los Inventarios consisten en desde puntos o líneas localizadas en el bosque, desde las que se seleccionan árboles. El número de UM que pueden localizarse en un bosque, independiente de su superficie, es infinito, o indeterminado. De manera que no existe un marco poblacional definido, como en el MAS tradicional, donde el número de individuos (UM) de la población es finito = N.

Las UM probabilísticas empleadas en los inventarios producen valores que son totales por hectárea de un atributo x medido en todos los árboles seleccionados en ella:

$$Xha_i = \sum_{j=1}^{m} F_j x_j$$

Donde, m es el número de árboles seleccionados en una parcela, x_j es el atributo medido en una arbol seleccionado y F_j es el factor de expansión, es decir, el número de árboles por hectárea que representa cada arbol medido en la parcela, que es función de la probabilidad de selección que se le ha asignado.

Promedio por hectárea

Los estimadores puntuales Xha_i tienen la propiedad de ser estimadores puntuales insesgados del total por hectárea. Si se localizan varias (n) UM probabilísticas, el promedio de las estimaciones puntuales es una estimación insesgada del total poblacional Xha = X/A:

$$\hat{X}ha = \overline{Xha} = \frac{1}{n} * \sum_{i=1}^{n} Xha_i$$

En la expresión anterior, Xha_i puede corresponder al valor agregado de una variable dasométrica cualquiera o al valor de la misma para una determinada clase de diámetro, clase de especie, etc. Este estimador permite, pues determinar el valor de una VE desagregada como las que integran las Tablas de Rodal y de Existencia

Error de estimación del promedio por hectárea

Una muestra de UM probabilísticas permite, igual como en el MAS tradicional, estimar la varianza de los estimadores $\hat{X}ha$:

$$S^2_{\hat{X}ha} = S^2_{\overline{Xha}} = \frac{s^2}{n} * (1 - f)$$

Donde:

s^2 **es la varianza muestral entre los valores** Xha **determinados en cada u.m.:**

$$s^2 = \frac{\sum_{i=1}^{n} (Xha_i - \overline{Xha})^2}{(n-1)} = \frac{\sum^{n} Xha^2_{i_i} - \left(\sum^{n} Xha_i \right)^2 / n}{n-1}$$

f **es la fracción muestral que normalmente es irrelevante y se omite. Sin embargo, en rodales pequeños muestreados muy intensamente su omisión puede generar un**

sobredimensionamiento del error de estimación. Su valor puede determinarse aproximadamente con la siguiente expresión:

$$f = \left(\sum_{i=1}^{n} \frac{1}{m} \sum_{j=1}^{m_i} a_j\right) / A \tag{14}$$

Donde a_j es el área de selección del árbol j, según el tipo de UM empleado, y A es la superficie del rodal. Cuando se emplean parcelas convencionales, la expresión de **f** es más simple:

$$f = \frac{n * a}{A}$$

Una expresión relativa (porcentual) del error de estimación es más práctica:

$$S\% = 100 * \frac{S_{\overline{Xha}}}{\overline{Xha}} = \frac{s\%}{\sqrt{n}} * \sqrt{1 - f}$$

Donde s% es el coeficiente de variación de las UM:

$$s\% = 100 * \frac{s}{\hat{X}ha}$$

Total Poblacional y su error

Expresiones del total poblacional y su correspondiente error de estimación, son las siguientes:

$$\hat{X} = A * \overline{Xha}$$

$$S_{\hat{X}} = A * S_{\overline{Xha}}$$

Error muestral

El error muestral, es decir, la máxima diferencia probable entre el valor estimado del promedio o del total y los correspondientes valores reales, para un nivel de confianza 1-α. En términos relativos (%) es el siguiente:

$$E\% = t_{1-\alpha/2}^{n-1} * S\%$$

Donde **t** es el valor tabulado de la distribución de t para un nivel de incertidumbre α y para n-1grados de libertad. Igualmente pueden obtenerse expresiones para el error muestral del promedio por hectárea y del total, al multiplicar el valor de **t** por la correspondiente expresión del error de estimación

Limites confidenciales

Los limites confidenciales representan el rango en que se encuentra el valor real, con una determinada probabilidad. Se establecen al sumar y restar el error muestral al valor (promedio o total) estimado. De esta manera se define un rango dentro del cual se encuentra el verdadero valor poblacional con un nivel de confianza de 1-α:

$$\text{L.C.(Xha)}= \overline{Xha} *(1 \pm \text{E\%}/100)$$

$$\text{L.C.(X)}= \hat{X} * (1 \pm \text{E\%}/100)$$

Es interesante destacar que si se usan UM convencionales, los estimadores descritos son idénticos a los estimadores correspondientes al diseño muestral MAS tradicional. Las expresiones de los estimadores clásicos del **MAS** no son aplicables, sin embargo, cuando se emplean otros tipos de UM. probabilísticas que no sean parcelas convencionales.

Ejercicios

En la Tabla 3 se presenta a modo de ejemplo, una estimación realizada sobre la **Población I** Dadas las características de esta población, se pueden emplear en ella, tanto los estimadores MAS tradicional, como los Alternativos. (figura 4). Esta población representa a un rodal de 40 ha, dividido en 400 unidades (individuos o UM), de 0,1 ha cada uno. Sobre una población (UI) de este tipo es posible realizar un proceso de selección muestral de tipo **MAS** tradicional, ya que el numero de individuos en la población es finito y se encuentran perfectamente identificables al momento de efectuar una selección muestral aleatoria.

Tabla 3. Ejemplo de selección aleatoria de 10 u.m. sobre población de juego I y empleo de estimadores MAS Alternativos

Volumen por hectárea(Xha_i): (200,88,260,77,141,106,59,171,159,12)			
Parámetro	**Símbolo**	**Valor**	**Observaciones**
Superficie parcela, ha	**A**	0,1	Parcelas convencionales (sup. Fija)
Superficie de la UI, ha	**A**	40	
Tamaño muestral	**N**	10	
Fracción muestral	**F**	0,025	$f = \frac{n*a}{A} =$
Promedio por hectárea	$\hat{X}ha$	127,3	$\hat{X}ha = \overline{Xha} = \frac{1}{n} * \sum_{i=1}^{n} Xha_i$
Total estimado del rodal	$\hat{X}$	5092	$\hat{X} = A * \overline{Xha}$
Varianza muestral	$\mathbf{s^2}$	73,40	$s^2 = \frac{\sum_{i=1}^{n} (Xha_i - \overline{Xha})^2}{(n-1)}$
Coeficiente de variación, %	**S%**	57,7	$s\% = (s / \hat{X}) * 100$
Error de estimación	$S_{\overline{Xha}}$	21,10	$S^2_{\hat{X}ha} = S^2_{\overline{Xha}} = \frac{s^2}{n} * (1-f)$
Error de estimación, %	$S\%$	15,79	$S\% = 100 * \frac{S_{\overline{Xha}}}{\overline{Xha}} = \frac{s\%}{\sqrt{n}} * \sqrt{1-f}$

Valor de t	$t_{1/\alpha/2}^{n-1}$	2,26	Para α = 0,05 y 9 grados de libertad
Error muestral, %	**E%**	35,69	$E\% = t_{1-\alpha/2}^{n-1} * S\%$
L.C. del promedio, inferior	**L.C.(Xha)**	81,87	L.C.(Xha)= $\overline{Xha}$ *(1 ± E%/100)
, superior		172,73	
L.C. del total, inferior	**L.C.(X)**	3275	L.C.(X)= $\hat{X}$ * (1 ± E%/100)
, superior		6909	

Los limites confidenciales presentados en el cuadro 3 (3275 - 6909) indica que existe sólo una probabilidad de un 5% (α=0,05) que el valor real total se encuentre fuera de los limites establecidos.

Tamaño de la muestra

En términos estadísticos, el tamaño de la muestra depende de la variabilidad entre UM, del error muestral máximo admisible y del nivel de incertidumbre (α) tolerado. Una expresión simple que permite estimar el tamaño muestral es la siguiente:

$$n = \frac{t^2 * s\%^2}{E\%^2}$$

Esta expresión resulta al despejar **n** en la función de Error muestral, asumiendo que la fracción muestral **f** es irrelevante. El valor de t en la expresión debe ser el que corresponde al n resultante, dado el nivel de confianza (1-α) adoptado. Esto hace que muchas veces sea necesario iterar en el calculo de n hasta lograr la igualdad indicada

Si resultara que, dado el tamaño de la muestra, la fracción muestral fuese relevante, es necesario ajustar el valor obtenido de **n** en la siguiente forma:

$$n_a = \frac{n}{(1+f)}$$

La determinación de **n** supone conocida para la UI, la variabilidad entre UM (**s%**). El mejor medio para estimar la variabilidad de la Unidad es recurrir a la experiencia de otros inventarios realizados en situaciones similares. El realizar muestreos previos no es una buena opción, pues la variabilidad que afecta a la estimación de s% con muestras pequeñas es muy grande.

Tamaño muestral para una superficie de referencia diferente a la de la UI

Hay ocasiones en que un determinado error muestral máximo es requerido para una cierta superficie boscosa que puede ser mayor o menor que la de la UI en cuestión. Esta superficie puede corresponder, por ejemplo, a la del abastecimiento anual requerido por las instalaciones industriales de una empresa (**A_0**).El tamaño muestral para una UI de superficie A se puede aproximar en la siguiente forma:

$$n = \frac{A}{A_0} * \frac{t^2 s\%^2}{E\%^2}$$

Tamaño muestral para un Marco presupuestario restringido

A veces el factor restrictivo para dimensionar el tamaño de la muestra no es un error máximo admisible, sino el presupuesto disponible para realizar el inventario. Si el prosupuesto es P, los costos fijos (diseño, capacitación, control, etc) son C_0 y el costo variable por UM es **c,** entonces el tamaño muestral posible será:

$$n = \frac{P - C_0}{c}$$

Relación Tamaño-Variabilidad

La variabilidad entre UM en una UI se expresa comúnmente mediante el coeficiente de variación: $s\% = 100 * \frac{s}{\hat{X}ha}$ =cv%

La variabilidad depende de características estructurales de la UI y del tamaño de las UM que se emplean. Mientras más grandes son las UM menor es la variabilidad. La tasa de reducción del cv% al aumentar el tamaño de las parcelas es variable, y depende de características estructurales de la población.

Relación Tamaño-Variabilidad en un Bosque Aleatorio

El llamado Bosque Aleatorio se caracteriza básicamente por dos elementos:

- La ubicación de cualquier árbol en él es el resultado de un proceso aleatorio, equiprobabilístico e independiente. La ubicación de un árbol es totalmente independiente de la ubicación de cualquier otro árbol del bosque.
  El proceso que da origen a un bosque aleatorio se llama también proceso de Poisson, porque la cuenta de árboles en parcelas de superficie fija distribuidas al azar sigue la distribución de Poisson:

  $$p(x) = \frac{e^{\mu} \mu^x}{x!}$$

  Donde: p(x) es la probabilidad de contar x árboles en una parcela que posee μ árboles en promedio. Así, para un bosque que contiene N árboles por hectárea y se emplean parcelas de a hectáreas, $\mu = a \cdot N$
- La dimensión de cualquier atributo de un árbol de la población es independiente de la dimensión de sus vecinos, cualquiera sea su proximidad. Esto quiere decir que la correlación entre dimensiones de los árboles es nula, independiente de su distanciamiento.

En Prodan et al(1997) de demuestra que para un bosque aleatorio, el producto del tamaño de una parcela por su variabilidad es constante:

$$C\%_i^2 \cdot a_i = C\%_j^2 \cdot a_j = cte.$$

Esta relación permite conocer para un bosque aleatorio, el coeficiente de variación que se dará para un determinado tamaño de parcela (a_i) si se conoce el C% asociado a parcelas de un tamaño a_j.:

$$C\%_i = C\%_j \left(\frac{a_j}{a_i} \right)^{0.5}$$

Por ejemplo, si empleando parcelas de 500 m^2 el coeficiente de variación es 25%, para parcelas de 1000m^2 este será:

$$C\%_i = 25\left(\frac{500}{1000}\right)^{0.5} = 17.7\%$$

Relación tamaño-Variabilidad para cualquier bosque

La expresión anterior puede formularse también del siguiente modo:

$$C\%_i=(C\%j*ajk)a_i^{-k} = ba_i^{-k}$$

Donde: **b** es una constante determinada por el producto de cualquier tamaño de parcela por su correspondiente coeficiente de variación y

k es un exponente que vale 0.5 para un bosque aleatorio

Numerosos estudios han permitido observar una clara relación entre el coeficiente k y la estructura de los bosques. Bosques homogéneos, como son las plantaciones poseen valores k mayores que 0.5, mientras que los bosques de estructura irregular como son la generalidad de los bosques naturales, tienen valores menores que 0.5

Es posible desarrollar modelos de variabilidad para bosques con diferentes estructuras empleando una transformación logarítmica de la función que relaciona un tamaño de parcela con su coeficiente de variación:

$$Ln(C\%) = b + kln(a)$$

Esta relación se puede tomar como un modelo lineal, donde ln(C%) es la variable dependiente, ln(a) la variable independiente y b y k coeficientes, cuyos valores pueden determinarse mediante Mínimos Cuadrados Ordinarios.

Para determinar el valor k de un bosque se necesita información experimental consistente en los C% correspondientes a diferentes tamaños de parcelas, como los valores que se muestran en la tabla siguiente (Prodan et al,1997):

a, ha	.01	.02	.04	.16	.32	.64
C%	139	96.1	74.6	48.8	42.3	39.4

Al ajustar los datos al modelo mediante mínimos cuadrados, se obtienen los valores de b=3.425 y **k=-0.301**. El modelo que relaciona los coeficientes de variación con los tamaños de parcelas para este bosque sería el siguiente:

$$C\%_i = C\%_j\left(\frac{a_j}{a_i}\right)^{0.301}$$

En la Tabla 3 se comparan los C% reales con los estimados mediante el modelo ajustado para el mismo bosque y con los estimados para un bosque aleatorio, suponiendo conocido el C% correspondiente a parcelas de a==.04:

Tabla 3

Tamaño de parcelas: a, ha	C% real	C% estimado	
		K = 0.301	K=0.5(b. aleatorio)
0.01	139.0	113.2	149.2
0.02	96.1	91.9	105.5
0.04	74.6	74.6	74.6
0.16	48.8	49.1	37.3
0.32	42.3	39.9	26.4
0.64	39.4	32.4	18.6

Estimadores para unidades muestrales de tamaño variable

Se ha indicado que la condición básica para que las estimaciones realizadas mediante muestreo en una Unidad de Inventario sean insesgadas y con error de estimación conocido es que las Unidades Muestrales sean todas de **igual tamaño.** En algunas ocasiones, sin embargo, las UM localizadas en una UI resultan de tamaños desiguales:

- **Muestreo con parcelas de borde, truncadas (figura 5a)**
- **Muestreo con fajas de longitud variable (figura 5b)**
- **Muestreo con líneas de parcelas de longitud variable (figura 5c)**
- **Muestreo con parcelas elípticas de radio constante en la pendiente, luego de superficie variable al proyectarse sobre el plano horizontal.**

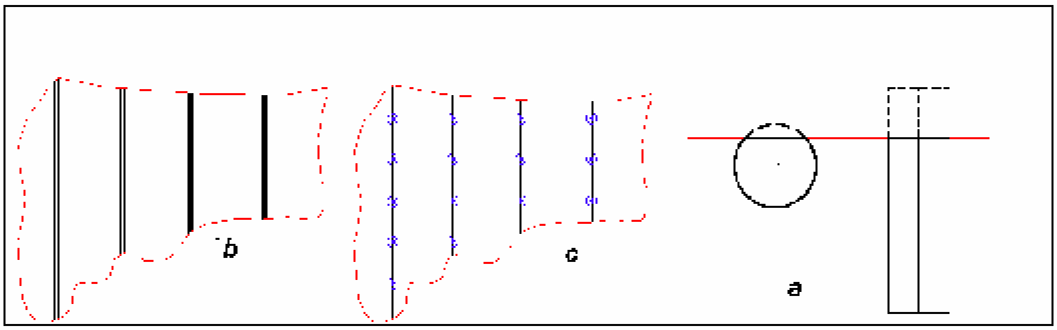


Figura 5

Cuando las parcelas medidas en un rodal resultan de dimensión variable, por cualquier causa, en lugar de los estimadores MAS deben emplearse los llamados **Estimadores de Razón de Medias (ERM).**

Simbología

n : Número de UM. Localizadas aleatoriamente en el rodal

x_i : Valor agregado de un atributo cualquiera (área basal, volumen,...) de todos los árboles medidos en la UM., sin expandir a valores por hectárea.

$$x_i = \sum_{j=1}^{m} x_j$$

x_i puede también corresponder a la suma de x para todos los árboles pertenecientes a una clase cualquiera k, en caso de crearse tablas de distribución de los parámetros (tablas de rodal y de existencia). Entonces, x_i podría ser el valor agregado del atributo x medido en todos los árboles pertenecientes a una determinada clase de especie y de diámetro k:

a_i : superficie de la UM. i, en hectáreas

$\bar{x}$: Promedio aritmético de los valores $\mathbf{x_i}$

$\bar{a}$: Promedio aritmético de la superficie de las parcelas $\mathbf{a_i}$

La simbología empleada para describir los estimadores es la misma que para el caso de UM de tamaño constante.

Total por hectárea:

$$\hat{X}ha = \frac{\overline{x}}{\overline{a}}$$

Total poblacional:

$$\hat{X} = A * \hat{X}ha$$

Error de estimación del promedio:

$$S^2_{\hat{X}ha} = \frac{\hat{X}ha^2 * (1-f)}{n(n-1)} \left(\frac{\sum^n x_i^2}{\overline{x}^2} + \frac{\sum^n a_i^2}{\overline{a}^2} - \frac{2\sum^n x_i a_i}{xa} \right)$$

La fracción muestral **f** puede aproximarse como en el MAS Alternativo:

$$f = \frac{\sum^n a_i}{A}$$

. Normalmente se asume igual a 0, pues en la práctica su valor es generalmente muy pequeño

Los restantes estimadores ($S_{\hat{X}}, S\%, E_{\hat{X}ha}, E_{\hat{X}}, E\%$ y limites confidenciales) se determinan en la misma forma que en el MAS Alternativo:

$$S\% = 100 * \frac{S_{\hat{X}ha}}{\hat{X}ha}$$

$$S_{\hat{X}} = A * S_{\hat{X}ha}$$

$$E(\hat{X}ha) = t^{n-1}_{1-\alpha/2} \cdot S_{\hat{X}ha}$$

$$E(\hat{X}) = A \cdot E(\hat{X}ha)$$

$$E\% = t^{n-1}_{1-\alpha/2} * S\%$$

$$\text{L.C.(Xha)} = \overline{Xha} * (1 \pm \text{E\%}/100)$$

$$\text{L.C.(X)} = \hat{X} * (1 \pm \text{E\%}/100)$$

Ejercicios

La Tabla4 presenta la Población de Juego II. Se trata de un rodal de 40 ha que se ha dividido en 40 franjas de un ancho igual a 31.32 m y longitud variable

Tabla 4.- Rodal de 40 ha dividido en 40 fajas de longitud variable

Faja N i	Superf. ha(zi)	Volumen m^3(yi)		Faja N i	Superf. ha(zi)	Volumen m^3(yi)		$\sum^N y_i = 5457$
1	0.4	78.9	*	21	0.5	117.9		$\sum^N z_i = 40$

2	0.5	114.8	*	22	0.7	163.2		$\sum^{N} z_i^2 = 47.88$
3	0.6	124.7		23	1.0	168.4		
4	0.7	156.1		24	0.8	146.8	*	$\sum^{N} y_i^2 = 868790.92$
5	0.8	149.0		25	0.7	117.2		
6	1.2	215.0		26	0.5	121.3	*	$\sum^{N} y_i z_i = 6074.89$
7	1.3	242.2		27	0.6	148.5	*	$\overline{Z} = 1.00$
8	1.4	246.9		28	0.7	126.6		
9	1.5	224.5		29	0.7	136.0		$\overline{Y} = 136.425$
10	1.6	256.9		30	1.0	177.3	*	$\sigma z = 0.4438$
11	1.8	168.6		31	1.0	155.4		$\sigma y = 55.7494$
12	1.7	154.9	*	32	1.0	87.3		$Sz = 0.4495$
13	1.5	124.8		33	1.3	155.0		$Sy = 56.4596$
14	1.2	96.1		34	1.5	166.9		$r = 0.6243$
15	1.0	82.9	*	35	1.5	149.7		$Syz = 15.843$
16	1.0	67.2		36	1.6	178.4		
17	0.8	67.9		37	1.7	191.4		
18	0.5	54.3	*	38	1.7	119.7		
19	0.3	29.4		39	1.0	80.7		
20	0.2	21.8		40	0.5	72.4	*	
				Total	40.0	5457.0		

Ejemplo

La Tabla 5 muestra un ejemplo de estimación con ERM sobre **Población II** (Tabla 4). Esta población, presentada en forma de tabla, corresponde a un rodal de forma irregular, dividido en fajas de ancho constante y longitud variable. Cada faja es considerada como un individuo, y para los efectos del muestreo, como una UM. Los individuos marcados con asterisco en la columna **Sel** son los que integran la muestra aleatoria de tamaño **n** =10.

Tabla 5. Resultados de un muestreo con ERM sobre población II

Parámetro	**Símbolo**	**Función**	**Valor**
Promedio por ha	$\hat{X}ha$		153,61
Total	$\hat{X}$		6144,5
Error de estimación	$S_{\hat{X}ha}$	**(*)**	19,19
Error de estimación, %	$S\%$		12,49
Error muestral, %	$E\%$	**(**)**	28,22
L.C. (prom./ha)			110,3 - 197,0
L.C.(Total)			4411 - 7878

(*) **Dado que en este caso particular, tanto N como n son conocidos, la**

Fracción muestral **f** corresponde al cuociente n/N

(**) El valor de **t** para **n**=10 y α=0,05 es 2,26

Es importante hacer algunas aclaraciones respecto a los estimadores ERM aplicados a inventarios:

1. Si bien estos estimadores son por principio sesgados, este sesgo prácticamente desaparece en muestras grandes
2. Cuando las UM son de tamaño constante, los estimadores MAS Alternativos y los estimadores ERM son iguales.
3. Existen estimadores semejantes a los ERM, llamados estimadores de media de razones (EMR). Ellos, a diferencia de lo ERM no tienden a ser insesgados al aumentar el tamaño de la muestra, por lo que su empleo es incorrecto. Muchas veces ellos se emplean inconscientemente, cuando UM de tamaño variable se expanden a valores por hectárea, dividiendo el valor agregado de la parcela ($\mathbf{x_i}$) por la superficie de la UM en hectáreas y aplicando estimadores MAS Alternativos sobre los valores individuales expandidos a totales por hectárea.

Dada la importancia que tienen los Estimaciones de Razón y de Regresión, este tema continua en el documento **Estimadores de Razón y de Regresión.**

Estimadores de proporciones

Los estimadores de proporciones se emplean en inventarios forestales para describir el estado del bosque en base a variables que pueden tomar solo dos valores posibles, como presencia o ausencia de un determinado atributo en los individuos, como especie, enfermedades, defectos, etc. Se pueden dar muchos ejemplos de variables de este tipo cuya descripción puede ser importante en inventarios con propósitos múltiples.

El siguiente ejemplo muestra cómo pueden emplearse los estimadores de proporción. Se trata de la regeneración de una determinada especie, observada en 20 parcelas de muestreo localizadas aleatoriamente en una UI de 40 hectáreas. Valores 1 o 0 registrados para cada parcela indican presencia o ausencia de dicha especie, sin considerar su abundancia. Los siguientes valores muestran el resultado de la regeneración observada en las 20 parcelas:

$\mathbf{x} = (0 , 0 , 1 , 0 , 1 , 1 , 0 , 0 , 1 , 0 , 1 , 1 , 0 , 1 , 0 , 0 , 1 , 0 , 0 , 0)$

$\mathbf{n_k}$ = 8 es el número de parcelas con presencia de regeneración de la especie k, y **n** = 20 el numero total de parcelas. $\mathbf{p_k = n_k/n}$ =8/20 = 0.4 es a la proporción de parcelas con presencia de regeneración de la especie en cuestión.

En una población integrada por N individuos, de los cuales N_k presentan el atributo k, se dice que la proporción de individuos en la población que presentan el atributo k es:

$$P_k = N_k/N$$

Puede demostrarse fácilmente que la proporción $\mathbf{p_k}$ basada en una muestra aleatoria, es un estimador insesgado de la proporción poblacional $\mathbf{P_k}$. La proporción muestr

Puede observarse que $\mathbf{p_k}$ corresponde también a la media aritmética de la variable **x**. Así también, la varianza muestral de x es:

$$s_x^{\ 2} = \frac{(0^2 + 0^2 + 1^2 + ...+ 0^2) - (0 + 0 + 1 + ...+ 0)^2 / n}{n-1} = \frac{n_x - n_x^{\ 2} / n}{n-1} = 0.253$$

El coeficiente de variación de x es el siguiente:

$$s\% = \frac{s_x}{\bar{x}} 100 = 126\%$$

Uno de los inconvenientes de los estimadores de proporciones es su alta variabilidad, especialmente cuando se trata de atributos escasamente representados

Conforme a los estimadores MAS, y puesto que p_k es estimador insesgado de P_k, el error de estimación de la media $\bar{x}$ es, igualmente el error de estimación de la proporción $\mathbf{p_k}$. Asumiendo que la fracción muestral **f** es irrelevante, se tiene

$$= 0.01263$$

El error de estimación de $\mathbf{p_k}$ en % es el siguiente:

$$S\% = 100*(S_p/p_k) = 28.1\ \%$$

Puesto que el área de la UI poblada con la especie k es $A_k = A*P_k$, una estimación MAS de A_k es la siguiente:

$$\hat{A}_k = A \cdot p_k = 40 * 0.4 = 16\ ha$$

De acuerdo a la forma en que se propagan los errores, el error de estimación de la superficie poblada por la especie k es:

$$S_{A_k} = S_{A \cdot p_k} = A \cdot S_{p_k} = 40 * 0.1124 = 4.5\ ha$$

Si la muestra (**n**) es suficientemente grande (**np** y **nq** son mayores que 5), el error muestral de la proporción $\mathbf{p_k}$ para un nivel de incertidumbre α es, según De Vries(1986), el siguiente:

$$E_{p_k} = t_{1-\alpha/2} * S_{p_k} + 1/(2n) = 0.245\ \ ;\ E\% = 100 * (E_{p_k} / p_k) = 61.2$$

los limites confidenciales para la proporción p y para la superficie estimada en base a ella, son los siguientes:

$$L.C.(p_k) = p_{k*}(1 \pm E\%/100) = (0.155 - 0.645)\ ;$$

$$L.C.(A_k) = A_{k*}(1 \pm E\%/100) = (6.2 - 25.8)\ ha.$$

Tamaño de la muestra n

Una expresión del error de estimación de la proporción en % es el siguiente:

$$S_p\% = \frac{S_p}{p_k} \cdot 100 = 100 \sqrt{\frac{1-p}{p(n-1)}} = 28.1$$

Si el error que afecta a la proporción $\mathbf{p_k}$ excediera las expectativas, sería preciso intensificar la muestra. De esta misma expresión puede derivarse un estimador del tamaño muestra. Para un error máximo admisible $S_p max\%$ (Ej: =5%), el n requerido es:

$$n = \frac{10000(1-p)}{p(S_p \max \%)^2} = 600$$

Estimación de Superficie mediante redes de puntos

Uno de los empleos más corrientes de los estimadores de proporciones es para estimar la superficies de diferentes Clases de Uso del suelo sobre Imágenes de Percepción Remota (IPR). El procedimiento consiste en distribuir una red aleatoria de puntos sobre una IPR cuya area total es A hectáreas. El área correspondiente a una clase de uso k, dentro de A es A_k y su proporción es $P_k = A_k/A$. Si se distribuyen n puntos al azar sobre la IPR dentro del area A y se cuentan los puntos ubicados dentro de la clase de uso k (n_k) , la proporción $p_k = n_k/n$ es un estimador insesgado de P_k, por lo tanto un estimador insesgado del area A_k y su correspondiente error de estimación son:

$$\hat{A}_k = A \cdot p_k$$

$$S_p = \sqrt{\frac{p(1-p)}{n-1}}$$

$$S_{A_k} = S_{A \cdot p_k} = A \cdot S_{p_k}$$

Estimación con red sistemática de puntos

En lugar de emplear una cantidad de puntos distribuidos aleatoriamente sobre la IPR, puede emplearse una red de puntos distribuidos sistemáticamente.

Si las Clases de Uso tienden a distribuirse aleatoriamente en el área A, no hay grandes ventajas al emplear una red sistemática, pues el error que afectaría entonces a la proporción estimada es el mismo que afecta a las estimaciones basadas en puntos aleatorios.

En cambio, si las clases de uso tienden a concentrarse dentro de A en subáreas de formas regulares, la red sistemática estima con errores más pequeños.

La regularidad del área de una determinada Clase de Uso k se mide y expresa mediante el llamado Cuociente perimetral C_k:

$$C_k = \frac{Per_k}{Per_0} = \frac{Per_k}{2\sqrt{\pi A_k}}$$

Donde: Per_k es el perímetro total de los elementos que constituyen la Clase de Uso k

Per_0 es el perímetro correspondiente a un círculo de área = A_k

La siguiente escala (Figura 6) permite formarse una idea de los valores C_k asociados con elementos con diferentes grados de regularidad:

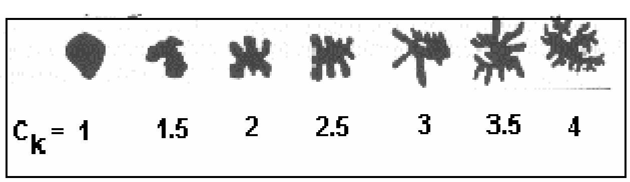


Figura 6

El error de estimación % para una red sistemática se determina en la siguiente forma:

$$S_p\% = 50.84 C_k^{0.5} n_k^{-0.75}$$

Ejercicio:

Sobre una área A = 50000 ha se distribuyen 300 puntos al azar y se cuentan 70 puntos en una clase de uso k. El área estimada para k es A_k = (60/300)*50000 = 10000 ha. El error de estimación (%) de A_k es

$$Sp\% = 100\sqrt{\frac{1-p}{p(n-1)}} = 100\sqrt{\frac{0.8}{0.2\cdot 299}} = 11.57\%$$

Si los mismos 300 puntos se hubieran distribuido en red sistemática cuadrada y se hubiesen contado los mismos 60 puntos en la Clase de Uso k, el error de estimación de la superficie de dicha clase, asumiendo que su cuociente perimetral es C_k=3, sería:

$$S_p\% = 50.84*3^{0.5}*60^{-0.75} = 4.1\%$$

Estimadores para grupos de Unidades

En muchas ocasiones hay interés en agrupar los resultados obtenidos para UI individuales con el fin de crear estimaciones de totales o totales por hectárea (o tablas de rodal y de existencia) para grupos de UI cuyo error de estimación sea también conocido.

También es frecuente que UI heterogéneas de subdividan en sub unidades o **estratos**, de manera que el error de estimación se reduzca al ser la variabilidad dentro de los estratos menor que la variabilidad en toda la UI.

En ambos casos se emplean los estimadores del diseño muestral llamado **Muestreo Estratificado**

Estratificación

Estratificar es dividir una población (bosque) en subpoblaciones (o estratos) integrados por individuos semejantes, de manera que la variabilidad dentro de ellos sea reducida. La clasificación en estratos se basa en general en imágenes de percepción remota. La asignación a estratos puede hacerse por delimitación o mediante clasificación de puntos distribuidos en una red sistemática.

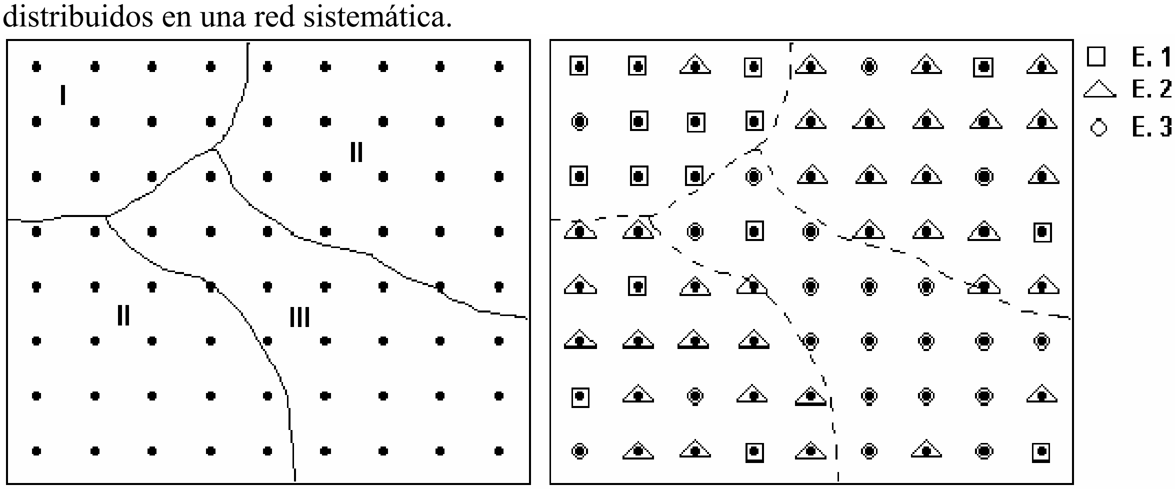


Figura 7

La figura 7 muestra los dos procedimientos de estratificación. Ambos procedimientos tienen ventajas y desventajas.
Los estratos delimitados tienen siempre una dimensión (superficie) que se asume exacta, pero la variación dentro de ellos es, en muchos casos, mayor. Esto ocurre, por ejemplo, en Clases de uso afectadas por una alta fragmentación, y donde se establece una superficie mínima delimitable que da origen a "islotes" de otras clases dentro de una Clase de Uso delimitada.
Los estratos resultantes de una clasificación puntual pueden ser bastante más homogéneos, pero no se conoce su magnitud exacta, debiendo emplearse diseños muestrales en fases. Además, los estratos no tienen una localización geográfica precisa, lo que dificulta la localización en terreno de los puntos elegidos como parcelas de terreno.

Estimadores

Simbología

Para cada unidad **h** con superficie **A_h** de las **L** UI que se agrupan o de los L estratos en que se subdivide una UI, se ha estimado el total $\hat{X}_h$ para el atributo de interés **x**. y su correspondiente error de estimación $S_{\hat{X}_h}$. Se asume que ambas estimaciones se han efectuado independientemente para cada rodal y en forma insesgada, es decir, sin errores sistemáticos de magnitud significativa. El tamaño muestral total n corresponde a la suma de los tamaños muestrales (n_h) empleados en las L UI. En cada UI la estimación puede basarse en un diseño muestral diferente.

Total de grupo

El total del atributo x para todas las unidades que integran el grupo(UI o estratos) y su correspondiente error de estimación son los siguientes:

$$\hat{X}_T = \sum_{h=1}^{L} \hat{X}_h \quad ; \quad S^2{}_{\hat{X}_T} = \sum_{h=1}^{L} S^2{}_{\hat{X}_h}$$

Promedio por hectárea

El promedio por hectárea y su correspondiente error de estimación son los siguientes:

$$\hat{X}ha_T = \frac{\hat{X}_T}{A_T} \quad ; \quad S_{\hat{X}ha_T} = \frac{S_{\hat{X}_T}}{A_T} \quad ; \quad A_T = \sum_{h=1}^{L} A_h$$

Error Muestral

El error muestral se calcula como en los casos anteriores, asumiendo una aproximación normal para los estimadores globales:

$$E(\hat{X}_T) = t_{1-\alpha/2}^{n-L} S_{\hat{X}_T} \quad ; \quad E\% = \frac{E(\hat{X}_T)}{\hat{X}_T} \cdot 100$$

Los grados de libertad que determinan el valor de "t" son aproximadamente igual a n-L, donde n= $\sum_{h=1}^{L} n_h$.

Tamaño Muestral

Partiendo de la siguiente expresión del Error Muestral es posible estimar el tamaño de la muestra en el Muestreo Estratificado:

$$E_{\hat{\bar{X}}}^{2} = t^{2} S_{\hat{\bar{X}}}^{2} = t^{2} \sum_{h=1}^{L} P_h^2 S_{\bar{X}_h}^{2} = t^{2} \sum_{h=1}^{L} P_h^2 \frac{s_x^2}{n_h}$$

Donde: P_h es la proporción del estrato h, es decir, A_h/A_T.

s_h^2 es la varianza entre unidades del estrato h, en valores absolutos

Antes de estimar n, es necesario especificar el criterio que se empleará para distribuir una muestra global de tamaño n en los L estratos. Los criterios de distribución más empleados son los siguientes:

a) **Distribución proporcional** a la superficie de los estratos: $n_h = n \cdot P_h$.

La expresión propia del Error Muestral es, entonces:

$$E_{\hat{\bar{X}}}^{2} = t^{2} \sum_{h=1}^{L} P_h^2 \frac{s_h^2}{n \cdot P_h}$$

..de la que se obtiene la expresión para estimar n:

$$n = \frac{t^2 \sum_{h=1}^{L} P_h s_h^2}{Em_{\hat{\bar{X}}}^2}$$

b) **Distribución óptima**, es decir, la que permite estimar con el mínimo error, dado un n:

$$n_h = n \cdot \frac{N_h s_h / \sqrt{c_h}}{\sum_{h=1}^{L} N_h s_h / \sqrt{c_h}}$$

Reemplazando en la expresión general del error muestral y despejando n, para costos constantes en todos los estratos, resulta:

$$n = \frac{t^2 (\sum_{h=1}^{L} P_h s_h)^2}{Em_{\hat{\bar{X}}}^2}$$

Tamaño muestral para un presupuesto fijo

Para un presupuesto P, con costo fijo (Diseño + capacitación + control + procesamiento) = C_0 y costo variable por parcela y por estrato c_h se tiene:

$$P = C_0 + \sum_{h=1}^{L} c_h n_h$$

Para una distribución proporcional al tamaño de los estratos, resulta:

$$C = C_0 + \sum_{h=1}^{L} c_h n \cdot P_h \qquad n = \frac{C - C_0}{\sum_{h=1}^{L} c_h P_h}$$

Ejercicios

La figura 8 presenta la Población de Juego I que se ha dividido en tres estratos

Estrato	Codigo	Ah, ha	Nh
1		10.0	100
2		19.8	198
3		10.2	102
Total		40.0	400

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	
1	130	153	153	112	200	106	100	147	118	165	0	0	12	0	35	0	18	0	0	24	
2	124	106	136	130	165	141	194	212	136	88	100	0	12	65	88	0	100	30	12	47	
3	177	165	136	124	171	106	82	177	147	165	118	82	47	6	88	12	30	0	0	24	A
4	165	112	124	118	153	118	224	136	118	159	141	65	35	24	0	30	30	53	53	30	
5	100	82	118	153	147	130	130	112	88	118	147	153	88	53	71	0	0	94	47	30	
6	224	247	217	230	130	259	277	100	147	171	200	171	118	141	82	59	71	6	0	0	
7	253	200	135	271	277	271	230	206	242	177	141	200	135	153	106	153	124	71	30	6	
8	212	277	265	212	206	171	289	259	183	247	194	277	183	165	88	106	118	136	53	71	B
9	224	283	247	300	100	318	277	306	177	200	177	271	141	71	124	71	188	171	159	94	
10	100	141	265	277	306	165	253	265	271	159	236	188	300	165	147	241	118	159	82	124	
11	277	330	253	218	177	353	330	253	171	194	241	177	177	118	88	106	118	188	77	165	
12	224	212	159	224	141	183	283	188	147	183	206	183	130	88	59	130	141	112	106	94	
13	271	318	200	271	218	253	260	200	147	259	253	77	165	242	153	194	106	224	59	141	C
14	277	277	206	236	230	230	294	165	294	212	259	159	94	124	212	100	159	124	218	200	
15	130	218	65	171	165	194	171	206	312	94	153	118	171	71	136	147	88	100	153	124	
16	218	130	118	130	82	171	147	124	177	183	159	94	124	212	100	159	124	100	82	71	
17	106	147	153	118	159	153	153	130	112	177	88	12	41	18	24	88	53	41	0	18	
18	130	200	194	100	141	165	153	147	177	194	106	35	0	18	0	0	35	30	41	35	D
19	77	165	159	159	183	118	124	124	94	159	71	0	100	18	6	6	0	0	0	30	
20	188	183	177	130	94	153	47	188	112	118	18	18	0	0	0	12	0	30	59	12	
			I					II					III					IV			

Figura 8

En cada estrato se han seleccionado parcelas aleatoriamente, con el resultado que se muestra en la Tabla 8

Tabla 8

Estrato	nh	Volumen por parcela, m^3/ha									
1	7	12	24	94	71	0	6	88			
2	15	136	171	141	130	141	200	141	159	183	77
		118	59	100	106	159					
3	8	135	283	100	171	218	260	200	247		

La Tabla 9 presenta los estimadores con y sin estratificación

Tabla 9

Pará-metro	Estimación sin estratificación	Estimación con estratificacion			
		Estrato 1	2	3	Total
N	400	100	198	102	400
n	30	7	15	8	30
P_h		0.250	0.495	0.255	1.000
$\hat{X}ha$	131.0	42.1	134.7	201.8	128.7
$\hat{X}$	5240	421	2667	2058	5146

s_h	72.9	40.7	38.4	63.3	
s%	55.6	96.6	28.5	31.4	
S_{Xtot}	512	148	188.1	219.3	324.8
S_{Xha}	12.8	14.8	9.50	21.5	8.12
S%	9.77	33.15	7.05	10.65	6.31
$t_{1-\alpha/2}^{n-L}$	2.05	2.45	2.14	2.36	2.05
E(Xtot)	1044	362.6	402.5	517.6	665.8
E%	19.94	86.13	15.09	25.15	12.94

Estimación de tamaño muestral para un error máximo admisible

El error muestral resultante fue de 12.94%. Suponiendo que el máximo admisible es 8%, ¿Cuál debería ser el tamaño muestral?

a) Para una distribución de las UM proporcional al tamaño de los estratos:

$$n = \frac{t^2 \sum_{h=1}^{L} P_h s_h^2}{Em_{\hat{X}}^2}$$

- El valor de t, para un nc=.95 y asumiendo que la muestra crecerá a un n=50 sería =2.01
- El error máximo se expresa en los mismos términos que la variabilidad por estrato, es decir, en m^3/ha: Em=0.08*128.7=10.3
- En el siguiente cuadro se calcula el resto de los elementos necesarios para calcular n:

Estrato	P_h	s_h	$P_h s_h^2$	$P_h s_h$
1	0.250	40.7	414.12	10.18
2	0.495	38.4	729.91	19.01
3	0.255	63.3	1021.76	16.14
Suma	1.000		2165.79	45.33
			(suma)^2	2054.31

Luego, n para una distribución proporcional:

$$n = \frac{2.01^2 \cdot 2165.79}{10.3^2} = 82$$

.. y para una distribución óptima:

$$n = \frac{2.01^2 \cdot 2054.31}{10.3^2} = 78$$

1. COCHRAN, W.G. 1977. Sampling techniques. John Wiley and Sons. N.Y. 428 p.
2. DE VRIES, P. 1986. Sampling theory for Forest Inventory. A teach Yourself course. Springer Velag Berlín-Heidelburg. 399 p.
3. HUSCH, B.; MILLER, CHARLES J.; BEERS, THOMAS W. 1982 Forest Mensuration 3ª Ed. John Wiley & Sons, 402 p.

4. Husch, B. , Beers, T.W. y Kershaw, J.A, 2003. Foresta Mensuration, 4ª Edición. John Wiley & Sons, Inc. 443 p
5. LANLY, J.P. 1981. Mnual of Forest Inventory With special reference to mixed tropical Forests. FAO Forestry Paper 27 , 200 p.

6. LOETSCH, F. And K.E. HALLER. 1964. Forest Inventory. Vol. I BLV-München Basel, Wien. 436 p.
7. LOETSCH, F., ZÖHRER, F. And K.E. HALLER. 1973. Forest Inventory Volume 2. BLV Verlaggesellschaft Munchen. 469 p.
8. PETERS, R.; M. JOBET y S. AGUIRRE. 1985. Compendio de tablas auxiliares para el manejo de plantaciones de pino insigne. Instituto Forestal, Manual Nº 14, 140 p.
9. PHILIP, M.S. 1994. Measuring Trees an Forests. CAB International. 310 p.
10. PRODAN, M.; 1968. Forest Biometrics Pergamon Press Bayerisch Landwirts-Chafts Verlag Gmbh, 447 p.
11. PRODAN, M.; PETERS, R.; COX, F. Y REAL, P. 1997. Mensura Forestal. GTZ-IICA. Costa Rica. 586 p.
12. RAJ, D. 1968. Sampling Theory. Mc Graw-Hill, N.4.
13. SCHREUDER, MT.; GREGOIRE, T.G. Y WOOD, G.B. 1995. Sampling methods for Multiresource Forest Inventory. John Wiley & Sons. Inc. 433 p.
14. VANCLAY, J. 1994. Modelling growth an yield: Applications to mixed Tropical Forest. CAB International.
15. Van Laar, Anthonie y Akca, Alparslan. 1997. Forest Mensuration. Cuvillier Verlag, Goettingen, Alemania (418 p). ISBN: 3-89588-874-5
16. ZOEHRER, F. 1980. Forstinventur. Ein Leitfaden für Studium und Praxis. Verlag Paul Parey. Hamburg. Berlín, 207 p.