LA REGRESION PONDERADA EN LA ELABORACION DE ECUACIONES DE VOLUMEN

21430

Víctor M. Barrena A. (1)

RESUMEN

El presente artículo muestra que utilizar la regresión ponderada (Mínimos Cuadrados Ponderados), es la mejor alternativa en la elaboración de ecuaciones de volumen, puesto que permite resolver las violaciones a las hipótesis de los Mínimos Cuadrados que presentan las poblaciones forestales. De esta manera se pueden determinar las inferencias estadísticas sobre la ecuación de volumen. Los dos corresponden a 200 árboles provenientes de la microrregión Iberia-Iñapari del departamento de Madre de Dios en la Amazonía peruana. El factor de ponderación a utilizar en la regresión ponderada debe ser inversamente proporcional a la variancia del volumen, variancia que es función del cuadrado del dap y de la altura.

SUMMARY

This paper shows that the use of Weighted Regression (Weighted Least Squares) is the best alternative in the elaboration of Volume Equations, because it allows to solve the violations to the Least Squares Hypothesis in forestry populations. Statistical inferences on the Volume Equations may be obtained. Data come from 200 trees from the Iberia-Iñapari micro-region. The weight factor used in the Weighted Regression must be inversely related to the variance of the volume; this variance is a function of the square of the dbh and the height.

---

(1) Profesor auxiliar, Facultad de Ciencias Forestales, UNALM

REVISTA FORESTAL DEL PERU 15(2): 21-28. 1988

INTRODUCCION

El volumen de las masas forestales es comúnmente estimado a partir de una ecuación de volumen. Estas ecuaciones generalmente se obtienen a partir de un análisis de regresión basado en el método de los Mínimos Cuadrados; pero las poblaciones de árboles violan las hipótesis de este método. Entonces el problema se soluciona transformando las variables o ponderando la ecuación de regresión.

El objetivo del presente artículo es mostrar que las ecuaciones de volumen obtenidas a partir de una regresión ponderada, basada en los Mínimos Cuadrados Ponderados, dan la mejor estimación del volumen de masas forestales.

REVISION DE LITERATURA

Los métodos estadísticos son los más utilizados para construir tablas de volumen, debido a su objetividad y, principalmente, al desarrollo de la informática en los últimos años, la cual permite simplificar y facilitar los cálculos respectivos (Bouchon, 1974; Caillez, 1980). Las ecuaciones obtenidas por estos métodos son cada vez más populares, pues para calcular el volumen de una masa forestal es tan sólo necesario una ecuación y ya no una tabla completa (Loetsch et al., 1973; Wenger, 1984).

El método estadísticos más utilizado es el análisis de regresión basado en el método de Mínimos Cuadrados (Caillez, 1980; FAO, 1981; Furnival, 1961); pero no necesariamente los resultados obtenidos por este método son los mejores (Cunia, 1964).

Para que la aplicación de este método sea correcta debe cumplirse con la hipótesis de los Mínimos Cuadrados (Bouchon, 1974; Cunia, 1964; Philip, 1983), que puede resumirse así: a) que la muestra sea obtenida al azar, b) que la población tenga una distribución normal y, c) que la variancia sea homogénea (homocedasticidad).

Los mismos autores indican que las poblaciones de árboles no cumplen con estas hipótesis. Munro (1964) dice que la variancia de volumen de los árboles grandes es mucho mayor que la de los árboles pequeños, porque depende del valor del diámetro a la altura del pecho (dap) elevado al cuadrado, multiplicado por la altura (h), y todo elevado al cuadrado. Observa además que esta característica puede acarrear inferencias estadísticas erróneas porque se incumple con la última hipótesis, como lo afirman Chatterjee y Price (1977).

Cunia (1964) afirma que en la estimación de volúmenes de árboles la homogeneidad de variancias es la hipótesis más importante; pero que, como ya hemos visto, no se cumple con ella. Para lograr la homogeneidad de variancias es necesario transformar las variables o ponderarlas aplicando la regresión ponderada basada en los Mínimos Cuadrados Ponderados (Bouchon, 1974; Cunia, 1964; Philip, 1983).

Pero al realizar la transformación logarítmica de las variables, los valores estimados por la regresión siempre serán sesgados, pues no pasan por el promedio de los datos observados (Bouchon, 1974). Cailliez (1980) añade que si la transformación es logarítmica, la estimación subestimará siempre el volumen.

Por otro lado, Munro (1964) concluye que los Mínimos Cuadrados Ponderados mejoran la precisión de la estimación del volumen, como lo señalan Cunia (1964), Furnival (1966) e IBDF (1983).

Daniel y Wood (1980) indican que el factor de ponderación a usarse en este método debe ser inversamente proporcional a las variancias de los valores por predecir. Por lo tanto, el factor de ponderación en la estimación de volúmenes de árboles debe ser proporcional a la inversa del dap al cuadrado por la altura (Munro, 1964; Bouchon, 1974; Cailliez, 1980; Cunia, 1964).

Cunia (1964) afirma que a pesar que la distribución de la población no sea normal, los coeficientes de la regresión ponderada son los mejores estimadores lineales, tanto en el sentido estadístico como en la estimación de la variancia de la regresión.

Pero es necesario indicar que si no interesa determinar la precisión con la que se estima el volumen de un rodal y sólo interesa obtener un buen ajuste, entonces ni la ponderación ni la transformación de variables son esenciales (Cailliez, 1980).

MATERIALES Y METODOS

Los árboles medidos provienen de la microrregión Iberia-Iñapari, departamento de Madre de Dios, cuya zona de vida es bosque húmedo-sub tropical en transición a bosque húmedo-tropical.

Los datos corresponden a 200 árboles de cualquier especie, escogidos al azar, que conforman una submuestra obtenida de la muestra con la que se elaboró la ecuación de volumen para esta zona (UNA, 1982).

Los dap han sido medidos con ayuda de una forcípula, mientras que los diámetros de las otras secciones y sus respectivas alturas han sido estimados con un relascopio de Bitterlich.

Los diámetos han sido tomados al nivel del suelo, al nivel del pecho, a 2,5 m de altura, al punto de copa y a la mitad de altura entre estos dos últimos niveles. Todos los diámetros han sido medidos con corteza.

Cada fuste ha sido dividido, según las mediciones hechas, en cinco secciones. El volumen de cada sección ha sido calculado utilizando la fórmula de Smalian; la suma de estos volúmenes fue considerada como el volumen total observado de cada fuste.

Como el objetivo del presente trabajo es verificar la conveniencia de utilizar la regresión ponderada en la estimación de los volúmenes de los árboles de la Amazonía peruana, se utilizará, entonces, un modelo simple de ecuación de volumen que permita obtener buenas estimaciones.

El modelo elegido, de acuerdo con la literatura consultada, denominado modelo 1, es el de la variable combinada:

$$V = b_o + b_1 dap^2 h \qquad (1)$$

Este modelo se ponderará por diversos factores de ponderación, lo que producirá otros modelos de ecuación de volumen. De esta manera, el modelo 2 estará ponderado por el factor $1/dap^2$; el modelo 3 por $1/dap^4$; el modelo 4 por $1/dap^2h$, y, por último, el modelo 5 por el factor $1/dap^4h^2$.

Como se menciona en la revisión de literatura, otra manera de lograr la homogeneidad de variancias es transformando las variables. En consecuencia, se utilizará un modelo logarítmico para probar todas las alternativas con la finalidad de lograr la homogeneidad deseada. El modelo escogido es el llamado de Schumacher:

$$\log V = b_o + b_1 \log dap + b_2 \log h \qquad (2)$$

que será el modelo 6.

Luego se comparará la bondad de las estimaciones de todos los modelos. Como las variables dependientes de estos modelos son diferentes y, por lo tanto, sus variancias están medidas en escalas diferentes, se utilizará el Indice de Furnival para realizar estas comparaciones (Furnival, 1961). El modelo que presente el menor valor en este Indice será el que estime mejor el volumen de los árboles de la submuestra.

Los cinco primeros modelos fueron procesados en el Centre de Traitement Informatique de l'Université Laval, Québec; se utilizó el paquete estadístico SAS.

El modelo logarítmico fue elaborado con ayuda de una IBM-PC, en la Facultad de Ciencias Forestales de la UNA La Molina.

RESULTADOS Y ANALISIS

Los valores de los parámetros b1 y del Cuadrado Medio del Error (CME) proveniente del análisis de variancia de cada regresión se presentan en el Cuadro 1.

Los valores del CME mostrados en este cuadro corresponden a ecuaciones de regresión que han sido ajustadas a variables dependientes diferentes (V, V/dap^2, V/dap^4, V/dap^2h, V/dap^4h^2, y $\log V$, respectivamente), por lo que no se pueden comparar entre ellos. Para compararlos se utilizó el Indice de Furnival cuya expresión general es la siguiente:

Cuadro 1. Valores de CME y de los parámetros b_i para cada modelo

Modelo	Factor de ponderación	CME	b_i		
			b_0	b_1	b_2
1	*	0.112	-0.036	5.553×10^{-5}	---
2	$1/dap^2$	3.057×10^{-5}	-0.042	5.554×10^{-5}	---
3	$1/dap^4$	1.161×10^{-8}	-0.031	5.468×10^{-5}	---
4	$1/dap^{2h}$	2.079×10^{-6}	-0.037	5.556×10^{-5}	---
5	$1/dap^4h^2$	5.802×10^{-11}	-0.014	5.430×10^{-5}	---
6	**	3.869×10^{-3}	-0.072	2.122	0.899

* Este modelo no fue ponderado

** Las variables fueron transformadas logarítmicamente

$$I = [f'(V)]^{-1} S \qquad (3)$$

donde: I = Indice de Furnival

f'(V) = Derivada de la variable dependiente con respecto al volumen

[f'(V)] = Media geométrica de esta derivada

S = Cuadrado Medio del Error

Según esta expresión, cuando no se han transformado las variables ni se ha ponderado la ecuación de regresión, entonces el Indice de Furnival será igual al CME.

$$I = S \qquad (4)$$

Pero si la ecuación de regresión es ponderada, entonces este Indice será igual a:

$$I = S \times \text{antilog} \left(\frac{1}{N} \log p \right) \qquad (5)$$

siendo p la inversa del factor de ponderación. Por último, si la ecuación es logarítmica, entonces:

$$I = S \times \log e^{-1} \quad \text{antilog} \left(\frac{1}{N} \log V_i \right) \qquad (6)$$

El cálculo de estos Indices se muestra en el Cuadro 2 donde N es igual a 200.

Estos resultados muestran claramente que los modelos ponderados (modelos 2 al 5) estiman mejor el volumen que un modelo no ponderado (modelo 1) o uno logarítmico (modelo 6), puesto que obtienen los menores valores del Indice de Furnival, es decir que sus estimaciones presentan el mejor error.

Además, si bien las diferencias entre los valores del Indice de Furnival de los modelos ponderados son pequeñas, se observa que los menores valores se obtienen cuando en el

Cuadro 2. **Cálculo del Indice de Furnival para cada modelo**

		1	2	3	4	5	6
S	(1)	0.112	3.057×10^{-5}	1.161×10^{-8}	2.079×10^{-6}	5.802×10^{-11}	3.869×10^{-3}
$\log (f'(V))^{-1}$	(2)	0	-146.652	-299.304	79.118	158.2361	23.755
(2)/N	(3)	0	0.733	-1.467	0.396	0.791	0.119
antilog (3)	(4)	1	0.185	0.034	2.487	6.183	1.315
							$x(\log e)^{-1}$
I		0.112	5.649×10^{-6}	3.964×10^{-10}	5.171×10^{-6}	3.587×10^{-10}	1.171×10^{-2}
(1) x (4)							

factor de ponderación se considera al dap al cuadrado multiplicado por la altura, y todo esto, a su vez, elevado al cuadrado; es decir, el cuadrado de la variancia del volumen.

CONCLUSIONES

1. La regresión ponderada brinda las mejores posibilidades en la estimación de volúmenes de masas forestales a partir de ecuaciones de volumen.

2. El factor de ponderación a utilizar debe ser proporcional al cuadrado de la inversa de la variancia del volumen. Si en el modelo el volumen es función del dap, entonces el dap al cuadrado intervendrá en el factor de ponderación; pero si el volumen es función del dap y la altura, en ese caso intervendrá el dap al cuadrado por la altura, todo elevado al cuadrado.

3. Se recomienda utilizar la regresión ponderada para elaborar ecuaciones de volumen si se quiere obtener, además de un buen ajuste, las inferencias estadísticas de la ecuación.

BIBLIOGRAFIA

BARRENA, V.; DANCE, J.; SAENZ, D. 1986. Metodología para la selección de ecuaciones de volumen. Rev. For. del Perú 13 (2): 3-12.

BOUCHON, J. 1974. Les Tarifes de Cubage. Nancy, France, Ecole National du Génie Rural des Eaux et des Forets. 57 p + anexos.

CAILLIEZ, F. 1980. Estimación del volumen forestal y predicción del rendimiento, con especial referencia a los trópicos. Vol. I. Estimación del Volumen. Estudios FAO Montes N°22/I Roma, 92 p.

CUNIA, T. 1964. Weighted Least Squares Method and Construction of Volume Tables. Forest Science 10 (2): 180-191.

CHATTERJEE, S.; PRICE, B. 1977. Regression Analysis by Example. New York, John Wiley and Sons. 228 p.

DANIEL, C.; WOOD, F. 1980. Fitting Equations to Data. New York, John Wiley and Sons. 458 p.

FAO. 1981. Manual de inventarios forestales, con especial referencia a los bosques tropicales heterogéneos. Estudios FAO: Montes N° 27. Roma. 200 p.

FURNIVAL, B. 1961. An Index for Comparing Equations Used in Constructing Volume Tables. Forest Science 7 (4): 337-341.

IBDF. 1983. Inventario Florestal Nacional; Florestas Nativas. Rio Grande do Sul, Brasil. 345 p.

LOETSCH, F.; ZOHRER, F.; HALLER H.E. 1973. Forest Inventory. Vol. II. Munchen, B.L.V. Verlagsgesellschaft. 469 p.

MUNRO, D. 1964. Weighted Least Squares Solutions Improve Precision of Tree Volume Estimates. Forest Chronicle. 40 (3): 400-401.

PHILIP, M. S. 1983. Measuring Trees and Forest; A Textbook writen for students in Africa. Tanzania, University of Dar-es-Salaam. 338 p.

UNA. 1982. Evaluación y plan de manejo de los recursos forestales de la microrregión Iberia-Iñapari. T-I. Lima, Dpto. de Manejo Forestal. 106 p.+ anexos.

WENGERS, K. F., Ed. 1984. Forestry Handbook. Society of American Forest. John Wiley and Sons. 1335 p.