METODOLOGÍA PARA LA SELECCIÓN DE ECUACIONES DE VOLUMEN

Víctor Barrena A.[1]
José Dancé C.[2]
Domingo Sáenz Y.[3]

RESUMEN

En el proceso de elaboración de una ecuación de volumen, generalmente se obtienen varias ecuaciones de regresión de las que se debe elegir una. En este artículo se presenta una metodología para seleccionar la mejor ecuación de volumen, utilizando para ello los valores del Cuadrado Medio del Error (CME) y si fuera necesario, los valores del Índice de Furnival (1) si las variables dependientes son diferentes. Los datos corresponden a 423 árboles que provienen de la Unidad Modelo de Manejo y Producción Forestal Dantas. La ecuación de volumen retenida corresponde al modelo:

$$\log V = b_o + b_1 \log D + b_2 \log H$$

Los resultados muestran que la metodología propuesta es la adecuada para elegir la mejor ecuación de volumen.

SUMMARY

Generally, in the construction of volume equations, several regression equations are obtained an just only one must be chosen. This paper shows a method for selecting the best volume equation using the Residual Mean Square (RMS) and, if it is necesary, the Furnival Index (1) if the dependant variables are different. Data comprises 423 trees from the Unidad Modelo de Manejo y Producción Dantas. The model of volume equation retain is:

$$\log V = b_0 + b_1 \log D + b_2 \log H$$

The results shows that the method proposed is adequate to choose the best volume equation

1. INTRODUCCION

Las "Ecuaciones de Volumen" son utilizadas para estimar el volumen de un árbol en función de alguna de sus características (diámetro a la altura de pecho, altura de fuste); ellas se obtienen por medio de un análisis de regresión. Este tipo de ecuaciones son actualmente más utilizadas que las tablas de volumen tradicionales.

La ventaja de utilizar ecuaciones de volumen en los inventarios forestales es evidente; ellas permiten obtener una estimación de volumen de toda una población de árboles a partir de medidas detalladas tomadas sobre un pequeño número de ellos.

Pero, un conjunto inapropiado de ecuaciones de volumen, o más propiamente dicho, malas ecuaciones de volumen, pueden reducir significativamente la confiabilidad de los resultados. Por eso

[1] Docente Contratado. Departamento de Manejo Forestal. Universidad Nacional Agraria La Molina..
[2] Profesor Principal. Departamento de Manejo Forestal. Universidad Nacional Agraria La Molina.
[3] Profesor Asociado. Departamento de Estadistica e Informática. Universidad Nacional Agraria La Molina

es importante, el seleccionar una ecuación de volumen adecuada que permita la mejor estimación de los volúmenes de los árboles inventariados.

El objetivo del presente artículo es, entonces, presentar una metodología para seleccionar la ecuación de volumen que mejor estime el volumen de los árboles. Paralelamente, se presentará el proceso seguido para seleccionar la ecuación de volumen que se utilizará en el inventario forestal de la primera unidad de manejo de la Unidad Modelo de Manejo y Producción Forestal Dantas.

II. REVISION DE LITERATURA

El trabajo de construcción de Tablas de Volumen consiste, según Philip (1983), de tres partes:

- Medición de los volúmenes individuales de los árboles seleccionados que constituyen una muestra representativa de la población.
- Establecimiento de las relaciones entré las mediciones tomadas sobre los árboles y sus volúmenes.
- Elección del mejor modelo y verificación de la precisión de la tabla de volumen elaborada,

Los procedimientos de construcción de Tablas de Volumen, es decir, los métodos que se emplean para elaborar una tabla de volumen se clasifican en: métodos directos, métodos gráficos y métodos estadísticos (Bouchon, 1974; Cailliez, 1980) siendo estos últimos los mayormente utilizados, pues el inconveniente que representaban los cálculos ha disminuido con el desarrollo de las computadoras (Cailliez, 1980).

Debido al creciente uso de las computadoras, las tablas de volumen como tales han perdido su importancia en los trabajos de inventariación forestal puesto que las ecuaciones de volumen, obtenidas por los métodos estadísticos, pueden ser incorporadas directamente en el programa de cálculo del volumen (Loetsch *et al*, 1973).

Dentro de los métodos estadísticos se tienen por ejemplo, el método de los mínimos cuadrados, algunos métodos del análisis multivariante no paramétrico y clasificación automática de datos, así como el análisis de regresión. Este último método permite la aplicación de la ecuación obtenida a los árboles (FAO, 1974).

Philip (1983) dice que la utilización de análisis de regresión en la construcción de ecuaciones de volumen presenta las siguientes ventajas:

- Proporciona un estimado de la precisión de la predicción o estimación
- Se cuenta con un método objetivo de elección entre diferentes modelos matemáticos
- Se necesitan pocos datos para alcanzar una precisión dada.

Si bien la intención del presente artículo no es la de mostrar en detalle las técnicas de regresión, es necesario conocer por lo menos las principales suposiciones en las que se basa los mínimos cuadrados para estimar los parámetros de la regresión. Según Philip (1983) éstas son:

1. Las unidades muéstrales (árboles) son seleccionados independientemente.
2. Los parámetros medidos sobre cada árbol son independientes.
3. La variancia del volumen de los árboles es constante e independiente con respecto a las variancias de las variables explicativas.

Además, en la estimación de los errores estándares y de los límites de confianza, se asume que:

4. Las mediciones son hechas sin error.
5. Las diferencias entre los volúmenes estimados y los medidos son distribuidas normalmente con una media de 0 y una variancia constante, estimada por el Cuadrado Medio del Error (CME) del Análisis de Variancia de la Regresión.

Según el mismo autor, los datos provenientes de árboles generalmente sólo cumplen con la primera suposición.

Bouchon (1974) por su parte, dice que se debe cumplir con las dos condiciones siguientes:

- Los errores de muestreo son independientes y se distribuyen normalmente.
- La variancia de la población es constante a lo largo de toda “la nube de puntos" correspondientes a los datos,

Pero añade que una población de árboles no cumple con el requisito de la homogeneidad de la variancia. Para corregir este defecto (heterocedasticidad) es recomendable transformar las variables (por ejemplo una transformación logarítmica) o ponderar las variables para aplicar la regresión ponderada (mínimos cuadrados ponderados) (Bouchon, 1974; IBDF, 1983, Philip, 1983).

En cuanto a determinar que variables utilizar y de qué manera ellas intervienen en la estimación del volumen, o sea dicho de otro modo, que modelos de ecuación de volumen probar, se recomienda primero, hacer una revisión de literatura para compilar los modelos que establezcan las relaciones más prometedoras entre las variables (Philip, 1983) porque de otra manera, se tendrían que probar todos los modelos estadísticos posibles para llegar a establecer la ecuación óptima (Caballero, 1972).

La elección de la mejor ecuación de volumen debe hacerse lo más objetivamente posible; para ello las ecuaciones deben clasificarse, según Philip (1983), de acuerdo a los siguientes criterios:

1. La bondad del ajuste de la ecuación a los datos, medida por el Coeficiente de Determinación (R^2).
2. El error estándar de la media.
3. El Cuadrado Medio del Error (CME).
4. El Índice de Furnival.

El R^2 se emplea si la ecuación de regresión describe un proceso; si la ecuación estima o predice valores, entonces se emplea el CME (Chatterjee y Price, 1977). Se utiliza el último criterio, el índice de Furnival, para comparar las ecuaciones que no tienen idéntica variable explicada ya que las variancias de ellas no se expresan en las mismas unidades (IBDF, 1983; Philip, 1983).

III. DATOS

Los datos corresponden a la muestra tomada para elaborar una tabla de volumen a utilizar en el inventario forestal de la primera unidad de manejo de la Unidad Demostrativa de Manejo y Producción Forestal Dantas. Estos datos fueron tomados en junio de 1983.

Se evaluaron 423 árboles de toda especie en los que se midieron, con ayuda del relascopio de Bitterlich, 5 diámetros y sus respectivas alturas. Los diámetros medidos fueron los siguientes: diámetro a 0.30 m sobre el nivel del suelo; diámetro a la altura del pecho (dap, a 1.30 m); diámetro a 2.30 m sobre el nivel del suelo; diámetro al punto de copa (altura de fuste) y el diámetro correspondiente al nivel equidistante a los dos últimos diámetros.

De esta manera, se dividió a cada fuste en 4 secciones y a cada una de ellas se les aplicó las fórmulas de Smalian y Newton para determinar el volumen individual de cada fuste. Este volumen medido fue utilizado, para ajustar los modelos.

IV. METODOLOGIA

Antes de decidir qué modelos de ecuación de volumen probar, se buscó en la literatura para seleccionar los modelos más prometedores (Caballero, 1972, IBDF, 1983, Loetsch et al, 1973). Los modelos seleccionados se presentan en el Cuadro 1.

CUADRO No. 1
MODELOS DE ECUACION DE VOLUMEN SELECCIONADOS

Modelos	Variables Independientes	Modelo	Denominación
1	D, H	$V= b_o+b_1 D+b_2DH+b_3D^2 +b_4H+b_5D^2H$	Comprensible (Meyer)
2	D, H	$V= b_o+b_1 D+b_2DH+b_3D^2 +b_4 D^2H$	Meyer modificada
3	D, H	$V= b_o+b_1 D^2 + b_2D^2H+b_3H^2+b_4 DH^2$	Naslund
4	D, H	$V= b_o+b_1 D^2 +b_2H+b_3D^2H$	Australiana
5	D, H	$V= b_o+ b_1 D^2 H$	De la variable combinada
6	D	$V=b_o+b_1D$	Berkhout
7	D	$V=b_o+b_1D_2$	Kopezky - Gehrhardt
8	D	$V=b_o+b_1D+b_2D^2$	Hohenald-Krem
9	logD, logH	$logV= b_o+b_1 logD+b_2logH$	Schumacher-Hall
10	logD, logH	$logV= b_o+b_1 logD^2H$	De la variable combinada logarítmica
11	log D	$loqV= b_o+b_1 log D$	Husch

Estos modelos fueron ajustados a nuestros datos por medio de una variante del procedimiento Forward Selection Stepwise" llamada "Método Stepwise" en el que si bien, básicamente es el procedimiento "Forward Selection" es decir, en el que a cada paso se añade una variable, pero con la diferencia de que a lo largo del proceso, una variable ya introducida puede ser eliminada en cualquier paso (Chatterjee y Price, 1977).

Para determinar cuál es la ecuación de volumen que mejor estima el volumen se procederá en tres etapas. La primera consiste en determinar la mejor combinación de variables (la mejor ecuación de regresión) para cada modelo, ya que no necesariamente todo el conjunto de variables de cada modelo explica el volumen de la mejor manera. Para ello se empleará el Cuadrado Medio del Error (CME). La mejor ecuación es aquella cuyo valor de CME es el menor.

Como lo muestra el Cuadro 1, existen dos grupos de modelos; uno en el que la variable dependiente es el volumen y otro donde la variable dependiente es el logaritmo del volumen. Como quiera que no se puede comparar directamente ecuaciones de regresión con variables dependientes diferentes como en este caso, se procederá primero a determinar la mejor ecuación de volumen dentro de cada uno de los grupos mencionados, y posteriormente, comparar las dos ecuaciones resultantes.

La segunda etapa consiste, entonces, en comparar dentro de cada uno de los dos grupos de ecuaciones obtenidas en el paso anterior. Para ello se compararán los valores del Cuadrado Medio del Error (CME) provenientes del Análisis de Variancia de cada regresión. La ecuación con el menor valor de CME explica mejor la variable dependiente.

De esta manera, se obtienen dos ecuaciones, una que explica el volumen y otra que explica el logaritmo del volumen. Ellas no se pueden comparar en base al CME porque las variantes corresponden a variables diferentes. Para compararlas se empleará el Índice de Furnival en la tercera etapa. El Índice de Furnival (1) es un índice que permite comparar las ecuaciones de volumen teniendo la ventaja de reflejar la dimensión de los residuales y además, las infracciones contra las condiciones de los mínimos cuadrados: linearidad, normalidad y homocedasticidad (Furnival, 1961).

La expresión general de este índice es la siguiente:

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

Donde:

I: Índice de Furnival
f' (V): Es la derivada de la variable dependiente con respecto al volumen
[f' (V)]: Media geométrica de esta derivada
S: Cuadrado Medio del Error.

Cuando la variable dependiente es alguna función del volumen, este índice puede ser considerado como un error estándar promedio transformado a unidades de volumen. En consecuencia, la ecuación que presente el menor valor del Índice de Furnival, será la mejor para estimar el volumen (Furnival, 1961). Por ejemplo, este índice ha sido utilizado en la determinación de la ecuación de volumen en los trabajos del Inventario Forestal del Estado Río Grande do Sul, Brasil (IBDF, 1983).

V. RESULTADOS Y ANALISIS

El procesamiento de los datos se realizó en el Centro de Procesamiento de Datos de la UNA (CEPDA). Para ajustar los modelos se utilizó el procedimiento Stepwise del paquete estadístico SAS.

Los valores del coeficiente de determinación (R^2) del CME correspondientes a la ecuación obtenida en cada paso para cada modelo, así como los valores de p (número de parámetros b a calcular en cada ecuación de regresión), se presentan en el Cuadro 2.

En la primera etapa del proceso de selección se ha utilizado el valor del CME para encontrar la mejor ecuación de regresión para cada modelo. La ecuación con el menor valor de CME, será la mejor.

Si el modelo tiene sólo una variable explicativa, se ajustó el modelo a los datos y la ecuación resultante será el resultado en esta etapa del proceso.

Los valores de R^2 y del CME de cada ecuación de regresión seleccionada para cada modelo se presentan en el Cuadro 3. Analizando los Cuadros 2 y 3 podemos decir que la mejor ecuación de regresión no es necesariamente la que corresponde al modelo completo. Es decir, no es aquella que considera un mayor número de variables independientes. Esto se puede observar claramente con los modelos 1, 2 y 3, donde las ecuaciones que mejor estiman el volumen no utilizan todas las variables del modelo correspondiente.

CUADRO No. 2

VALORES DEL R2 Y DEL CME DE LAS ECUACIONES DE REGRESION OBTENIDAS EN CADA PASO PARA CADA MODELO DE ECUACION DE VOLUMEN

Modelo No.	Paso	Variables	p	R^2	CME
	1	D^2H	2	0,793346	2,6228363
	2	D^2H, D	3	0,0091072	2,4286345
	3	D, D^2H, H	4	0,8179609	2,3215874
	4	D. D^2, D^2H, H		0,8203353	2,2968538
	4	D, D^2, DH, D^2 H	5	0,8220591	2,2748164
	5	D, D^2 , DH, D^2H, H	6	0,8220663	2,2802451
2	1	D^2H	2	0,793346	2,6228363
	2	D, D^2H	3	0,8091072	2,4286345
	3	D, DH, D^2H	4	0,815043	2,3587995
	3	D ' D^2, DH	4	0,8220021	2,2700482
	4	D, D^2, DH, D^2H	5	0,8220591	2,2748164
3	1	D^2H	2	0,793346	2,6228363
	2	D^2, D^2H	3	0,7992353	2,5542296
	2	D^2. DH^2	3	0,8173661	2,3235607
	3	D^2, DH^2, H^2	4	0,8175938	2,3262684
	4	D^2, D^2H, DH^2, H^2	5	.0.8179858	2,3268899
4	1	D^2H	2	0,793346	2,6228363
	2	D^2, D^2H	3	0,7992353	2,5542296
	2	D^2, H	3	0,809865	2,4189926
	3	D^2, D^2H, H	4	0,8169242	2,3348081
5		D^2H	2	0,793346	2,6228363
6		D	2	0,7184009	3,049358
7		D^2	2	0,7597402	3,049358
8	1	D^2	2	0,7597402	3,049358
	2	D, D^2	3	0,7657637	2,9800713
9	1	logi)	2	0,7921332	0,0403965
	2	logD, logH	3	0,8714873	0,0250351
10		logD^2, H	2	0,8704196	0,0251824
11		logD	2	0,7921332	0,0403965

CUADRO No. 3

VALORES DEL R^2 Y DEL CME DE LAS MEJORES ECUACIONES DE REGRESION PARA CADA MOELO DE ECUACION DE VOLUMEN

Modelo	Variables	p	R^2	CME
1	D, D^2 , DH, D^2 H	5	0,82205909	2,27481644
2	D, D^2 DH	4	0,82200213	2,27004819
3	D^2 , DH^2	3	0,81736605	2,32356068
4	D^2 , D^2 H, H	4	0,8169242	2,33480814
5	D^2H	2	0,79334596	2,62283631
6	D	2	0,71840087	3,57403317
7	D^2	2	0,75974019	3,049358
8	D, D^2	13	0,76576373	2,98007129
9	logD, logH	3	0,87148731	0,0250351
10	logD, logH	2	0,87041959	0,02518242
11	logD	2	0,79213315	0,04039646

Se aprecia además, que el valor de R^2 correspondiente a las ecuaciones seleccionadas, no es tampoco el mayor para cada modelo. Esto nos demuestra que si el criterio de selección de las ecuaciones de regresión es el R^2 podemos incurrir en un error. Es lógico que el R^2 aumente con el número de variables independientes utilizadas puesto que la variabilidad de la variable dependiente es explicada entonces por un mayor número de variables; pero esto no quiere decir que la estimación de la variable dependiente sea mejor.

Como ya hemos señalado, contamos con dos grupos de modelos, uno en donde la variable dependiente es el volumen (modelos 1 al 8) y otro en donde es el logaritmo del volumen (modelos 9 al 1l). En el siguiente paso se ha seleccionado la mejor ecuación de regresión de cada grupo comparando los valores del CME de las ecuaciones. La ecuación que estime mejor a la variable dependiente será aquella que posea el menor CME.

Como resultado de estas comparaciones se tiene que la ecuación del modelo 2 es la ecuación que explica mejor al volumen, ella utiliza como variables independientes al Dap, Dap^2, Dap x H y Dap^2 x H. Por otro lado, la ecuación del modelo 9 es la que explica mejor al logaritmo del volumen utilizando como variables explicativas al log Dap y al log H. Los valores de R2 y del CME de cada una de estas ecuaciones se muestran en el Cuadro 4.

CUADRO No. 4
VALORES DE R 2 Y DEL CME DE LAS MEJORES ECUACIONES DE VOLUMEN DE CADA GRUPO DE MODELOS

Modelo No.	Variables	R 2	CME
2	D, D 2 DH, D 2 H	0,82200213	2,27004819
9	logD, logH	0,87148731	0,02503510

Estas ecuaciones de regresión no se pueden comparar directamente porque las variables dependientes son diferentes; por ello utilizaremos el Índice de Furnival (1) para compararlas.

Para el cálculo de este Índice se requiere de la derivada de la variable dependiente con respecto al volumen. Como en la ecuación del modelo 2 la variable dependiente es el Volumen entonces el índice de Furnival para esta ecuación será:

$$I = S$$

Es decir, será igual al CME correspondiente a esta ecuación. Mientras que para la ecuación del modelo 9, cuya variable dependiente es el logaritmo del volumen (log V) el Índice de Furnival, de acuerdo a la expresión (1) se determinará de la siguiente manera:

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

En donde:

$$f'(\log V) = \frac{\log e}{V}$$

Entonces sus medias geométricas, también serán iguales:

$$[f'(\log V\ 1 = \frac{[\log e]}{V} \qquad (2)$$

Desarrollando tenemos que:

$$\frac{[\mathrm{Log}\ e]}{V} = \log e \times \frac{1}{n\sqrt{V_1 \times V_2 \times V_3 \ldots \times V_n}}$$

En donde la raíz n de esta expresión puede ser resuelta por medio de logaritmos, aplicándole posteriormente el antilogaritmo, por lo que esta última expresión puede ser escrita de la siguiente manera:

$$\frac{[\mathrm{Log}\ e]}{V} = \frac{\log e}{\mathrm{antilog}\left(\frac{1}{n} \sum \log V_i\right)}$$

Entonces reemplazando en la expresión (2) tenemos que:

$$[f'(\mathrm{Log}\ V)]^{-1} = \frac{\mathrm{antilog}\left(\frac{1}{n} \sum \log V_i\right)}{\mathrm{Log}\ e}$$

En consecuencia, el Índice de Furnival para esta ecuación, se obtiene a partir de la siguiente expresión:

$$I = S \times \log e^{-1} \times \mathrm{antilog}\left(\frac{1}{n} \sum \log V_i\right)$$

El cálculo de este índice se presenta en el Cuadro 5.

CUADRO No. 5 CALCULO DEL INDICE DE FURNIVAL

	Modelo No. 2	Modelo No. 9
S (1)	2,27004819	0,02503510
E log (f (V) $^{-1}$ (2)	0	86,66868
(2) / n (3)	0	0,208338
antilog (3) (4)	1	1,615616 (x [log e]-1)
[f'(V)] - ^{1} (5)	1	3,720093
I ((1) x (5))	2,27004819	0,093133

CUADRO No. 6
ANALISIS DE VARIANCIA DE LA ECUACION DE REGRESIONCORRESPONDIENTE AL MODELO 9 .

	GL	SC	cm	F
Regresión	2	70,45511126	35,22755563	1407,13
Error	415	10,38956733	0,02503510	
Total	417	80,84467861		

Este resultado muestra que la ecuación de volumen correspondiente al modelo 9, que utiliza el log V como variable dependiente, es la más apropiada para estimar el volumen en el inventario forestal de la primera unidad de manejo de la Unidad Modelo de Manejo y de Producción Forestal Dantas puesto que tiene el menor valor deL Índice de Furnival. Este modelo es el siguiente:

$$\log V = b_o + b_1 \log Dap + b_2 \log H$$

El Análisis de Variancia de este modelo se presenta en el Cuadro 6 y el valor de los coeficientes de la ecuación en el Cuadro 7.

CUADRO No. 7

COEFICIENTES Y SIGNIFICANCIA DE LOS MISMOS, DE LA ECUACION DE VOLUMEN CORRESPONDIENTE AL MODELO 9

	bi	Error Estándar	F
Intercepto	-0.40827712		
logD	1.66665873	0.03804032	1919.57
logH	0.95942173	0.05993406	256.25

VI. CONCLUSIONES Y RECOMENDACIONES

- La ecuación de volumen que mejor estima el volumen de los árboles de la primera unidad de manejo de la Unidad Modelo de Manejo y Producción Forestal Dantas es la siguiente:

$$\log V = -0.40827712 + 1.66665873 \log Dap + 0.95942173 \log H$$

- Para respetar las condiciones de linearidad, normalidad y de homocedasticidad de los residuales de las ecuaciones de regresión utilizadas para estimar el volumen de árboles, se recomienda realizar transformación de variables o ponderar la ecuación de regresión.

Se recomienda utilizar el Índice de Furnival si se quiere comparar ecuaciones de volumen con variables dependientes que no se expresen en las mismas unidades.

VI. BIBLIOGRAFIA

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

2. CABALLERO, M. 1972. Tablas y tarifas de volúmenes. México. Dirección General de¡ Inventario Nacional Forestal. Nota I.N.F. 4.2- 1, No. 7. 55 p.

3. CAILLIEZ, F. 1980. Estimación del volumen forestal y predicción de¡ rendimiento, con especial referencia a los trópicos. Estimación del volumen. FAO: Montes No. 22/1. 92 p. c. 1.

4. CHATTERJEE, S. y B. PRICE. 1977. Regression analysis by example. New York. John Wiley and Sons. 228 p.

5. FAO. 1981. Manual de inventarios forestales, con especial referencia a los bosques tropicales heterogéneos. FAO: Montes No. 27. 200 p.

6. FURNIVAL, G. 1961. An. Index for comparing equations used in constructing volume tables. Forest Science 7 (4): 337-341.

7. INSTITUTO BRASILEIRO DE DESENVOLVIMIENTO FLORESTAL

8. (IBDF). 1983. Inventario florestal nacional. Florestas nativas. Río Grande do Soul, Brasil. 345 p.

9. LOETSCH G., F., ZOEHRER y H.E. HALLER. 1973. Forest inventory. B.L.V. Verlagsgesellschaft-mbH, Munchen.469 p. v. 2.

10. PHILIP, M.S. 1983. Measuring trees and forest. A textbook writen for students in Africa. Tanzania. The Division of Forestry. University of Dar-es-Salaam. 338p.