# CAPÍTULO 6: REGRESIÓN y TABLAS y ECUACIONES DE VOLUMEN

por: Enrique Wabo.

## 1 MODELOS DE REGRESION

Como la construcción de Ecuaciones de Volumen utiliza la técnica de los Modelos de Regresión, comenzaremos repasando este tema.

### 1.1 INTRODUCCIÓN

Un Modelo de Regresión es un modelo matemático que expresa el valor esperado de una variable aleatoria Y, para valores distintos de una variable independiente o predictora X. La predicción se realiza a través de una función Y = f(X); la función f(X) recibe el nombre de regresión de Y sobre X. La función más simple es la de la línea recta, que da lugar a la Recta de Regresión de Y sobre X, por lo que el análisis se hará para este modelo. De todas formas, las deducciones y mecanismos son válidos para otros modelos.

Desde el punto de vista de la geometría analítica una recta se puede representar mediante la siguiente expresión:

$$\eta = A + BX \qquad (1)$$

siendo:

$\eta$ = el valor de la variable dependiente.

A = la ordenada al origen

B = pendiente de la recta.

Desde el punto de vista de la regresión, $\eta$ representa el valor esperado o medio de la variable de interés Y cuando la variable predictora X toma un valor particular xi (cuando X = xi), que es expresado como E(Y|xi). Así, la fórmula (1) toma la siguiente forma: E(Y|xi) = A + B xi, y que por regla general se expresa como:

$$Y = A + B\,x \qquad (2)$$

Definir la Recta de Regresión significa definir los valores de los coeficientes A y B del modelo para la población en estudio. Para ello recurrimos, a una muestra de Y con los valores asociados de X. Los valores estimados de A y B se indican con a y b, respectivamente. El modelo estimado es:

$$Y = a + b\,x \qquad (3)$$

Para llevar a cabo esta estimación, es necesario definir un estimador. El método más difundido es el Método de Mínimos Cuadrados. Otro método, más complejo, es el Método de Máxima Verosimilitud.

### 1.2 MÉTODO DE MÍNIMOS CUADRADOS (MMC)

El MCM establece que se deben tomar como estimación de los coeficientes del modelo, a partir de los datos de una muestra, los valores a y b que minimicen la suma S de los cuadrados de los desvíos con respecto a la media, usualmente llamada simplemente Suma de Cuadrados: $S = \Sigma(y - \hat{y})^2$.

$$S = \sum_{1}^{n} (Y - a - bX)^2 = \text{mínimo} \qquad (4)$$

Para su determinación se establecen las derivadas parciales de S con respecto a los términos a y b, que aquí actúan como variables, las que se igualan a cero:

$$\frac{\partial s}{\partial a} = -2\sum_{1}^{n} (Y - a - bX) = 0 \qquad (5a)$$

$$\frac{\partial s}{\partial b} = -2\sum_{1}^{n} (Y - a - bX)\,X = 0 \qquad (5b)$$

De estas dos derivadas parciales se obtiene un sistema de 2 ecuaciones, llamadas Ecuaciones Normales (no tienen relación alguna con la Distribución Normal). En el caso de la Recta, las Ecuaciones Normales, toman la siguiente forma:

$$an + b\sum_{1}^{n} x = \sum_{1}^{n} y \qquad (6a)$$

$$a\sum_{1}^{n} x + b\sum_{1}^{n} x^2 = \sum_{1}^{n} xy \qquad (6b)$$

El método de Máxima Verosimilitud "le presta" al MMC la fórmula de expresión de la varianza de la regresión si el modelo fuese exacto:

$$\sigma^2 = \frac{\sum (Y - A - BX)^2}{n} \qquad (7)$$

Si desarrollamos el sistema de ecuaciones obtenemos las soluciones para estimar los parámetros del modelo. Los datos básicos son los siguientes:

- S. de Cuadrados de los Desvíos de X (SCDX)
- S. de Cuadrados de los Desvíos de Y (SCDY)
- S. de los productos cruzados entre X e Y (SPXY)
- S. de Cuadrados de la Regresión (SCReg)
- S. de Cuadrados del Error o de los Desvíos de la Regresión (SCE)

La Figura 1 muestra el significado de la SC de la Regresión y del Error:

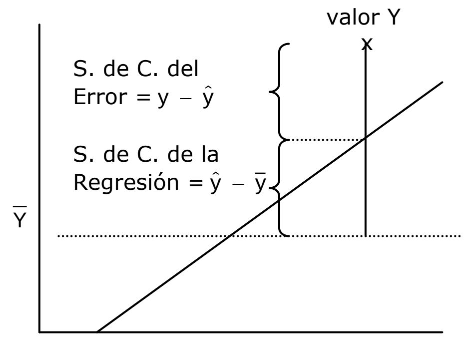



Figura 1

$$SCDX = \sum (x - \bar{x})^2 = \sum x^2 - \frac{\left(\sum x\right)^2}{n} \qquad (8)$$

$$SCDY = \sum (y - \bar{y})^2 = \sum y^2 - \frac{\left(\sum y\right)^2}{n} \qquad (9)$$

$$SPXY = \sum (x - \bar{x})(y - \bar{y}) = \sum xy - \frac{\sum x \sum y}{n} \qquad (10)$$

$$SCReg = \frac{(SPXY)^2}{SCDX} = b\ SPXY \qquad (11)$$

$$SCE = \sum (y - a - b)^2 = \sum (y - \hat{y})^2 \qquad (12)$$

$$= SCDY - SCReg = SCDY - \frac{(SPXY)^2}{SCDX}$$

Las estimaciones de las constantes son:

$$b = \frac{SPXY}{SCDX} \qquad (13)$$

$$a = \bar{y} - b\bar{x} \qquad (14)$$

El modelo visto es un modelo lineal, y en general cualquier modelo que puede ser expresado como polinomio o parte de un polinomio es un modelo lineal. Por ejemplo, el modelo:

$$Y = a + bX + cX^2 + dX^3 \qquad (15)$$

es un modelo lineal aunque su forma geométrica sea una curva.

Por el otro lado, existen los modelos no lineales, como por ejemplo:

$$V = aX^b \qquad (16)$$

Para aplicar el MMC en modelos no lineales, éstos deben ser linealizados y para ello recurrimos a los logaritmos. Así, la forma linealizada del modelo (16) es:

$$Ln\ (Y) = a' + b\ Ln\ (X) \qquad (17)$$

siendo a' = Ln (a). Una vez linealizado el modelo, el procedimiento es igual que con los modelos lineales en casi todos sus aspectos.

## 1.3 ANÁLISIS DEL COMPORTAMIENTO DEL MODELO

Una vez seleccionado el modelo y estimado sus parámetros debemos analizar su comportamiento. El objetivo del análisis es buscar elementos que por no ser satisfactorios nos habiliten para rechazar el modelo. Los más importantes son el Análisis de Varianza, el Coeficiente de Determinación ($R^2$) y el Análisis de los Residuales.

### 1.3.1 Análisis de Varianza

La regresión puede verse como una descomposición de la suma de cuadrados de los desvíos de los valores de Y respecto a su media (SCDY), en dos componentes.

Uno es la Suma de Cuadrados del Error, que representa la variación de Y independientemente de la variación de X; está expresada por la diferencia $(y - \hat{y})$. El otro es la Suma de Cuadrados de la Regresión, que representa la variabilidad de Y ligada a X; está expresada por la diferencia $(\hat{y} - \bar{y})$. En la Figura 1 se distingue cada uno de estos componentes.

Si cada una de estas suma de cuadrados se asocian con sus correspondientes grados de libertad, podemos determinar sus Cuadrados Medios y poner a prueba, mediante el cociente de cuadrados medios, la siguiente hipótesis: ¿varía significativamente Y cuando varía X? Este prueba es equivalente a la hipótesis si la pendiente B es distinta de cero, prueba que recurre a la variable "t" de Student. En el Cuadro 1 se muestran los componentes del ANOVA.

ANVA

| F. de V. | S. C. | g.l. | C.M. | F |
|---|---|---|---|---|
| Regresión | b SPXY | 1 | CMR | $\frac{CMR}{CME}$ |
| Error | SCDY – b SPXY | n-2 | CME | |
| Total | SCDY | n-1 | | |

El valor de F obtenido se compara con el valor tabular correspondiente.

El término CME representa la varianza de la regresión, que también se indica como $S_{yx}^2$; su raíz cuadrada es la desviación estándar de la regresión: $S_{yx}$.

### 1.3.2 Coeficiente de determinación ($R^2$) e Índice de Furnival

Este coeficiente es un indicador de la Bondad de Ajuste del modelo ensayado. Representa el por ciento de variación de Y que es explicado por el modelo; toma valores entre 0 y 1. Si fuese igual a 1 estaría indicando que todos los puntos de la muestra se ubican exactamente sobre la recta; o bien, que algunos puntos de la recta coinciden con los puntos de la muestra. Se lo define como:

$$R^2 = \frac{SCReg}{SCDY} \qquad (18)$$

Cuanto mayor es el valor de $R^2$ mejor es la *Bondad de Ajuste* del modelo. Según el tema con que tratemos es de esperar que su valor se encuentre en cierto rango; de no ser así podemos rechazar el modelo si no hay causa que justifique ese bajo valor, por ejemplo, una muestra muy chica.

El **Índice de Furnival** también es un indicador de la bondad de ajuste y se lo utiliza cuando se ensayan modelos no lineales. Esto ocurre porque los $R^2$ calculados mediante fórmulas, a mano o mediante computadora, no son directamente comparables. En el modelo lineal, las variables que llamamos X e Y son directamente las variables de interés X e Y. Pero en la versión linealizada de un modelo no lineal Y es en realidad el logaritmo de la variable Y original. De allí que no sean comparables. Para solucionar este problema Furnival desarrolló un indicador en reemplazo del $R^2$, llamado Índice de Furnival (IF).

Sea $f'(y)^{-1}$ la inversa de la derivada de la variable Y transformada, se define como Índice de Furnival a

$$IF = S_{yx} \left[ \prod_{1}^{n} f'(y)^{-1} \right]^{1/n} \quad (19)$$

El IF es el producto de la desviación estándar de la regresión y de la media geométrica de las inversas de la derivada de la variable Y transformada. Una fórmula más sencilla de usar es la siguiente:

$$IF = S_{yx} \text{ x Antilog} \left[ \frac{\sum_{1}^{n} \log \left[ f'(Y)^{-1} \right]}{n} \right] \quad (20)$$

En el siguiente Cuadro se indican las transformaciones más comunes y las inversas de su derivada:

| Transformación hecha f(y) | Inversa de la derivada $f'(y)^{-1}$ |
|---|---|
| Ninguna | 1 |
| log y | 2,3026 y |
| ln y | y |
| y/w | w |
| 1/y | $-(y^2)$ |
| $y^k$ | $1/(k \times y^{k-1})$ |

1.3.3 Análisis de residuales

El procedimiento consiste en representar en un gráfico: a) el comportamiento de los valores estimados de Y en función de los valores Y observados, y b) los residuos en función de los valores Y observados.

### 1.4 EJEMPLO

Determinaremos el modelo lineal de volumen en función del cuadrado del dap:

$$Y = a + b X$$

siendo Y el volumen sin corteza del fuste y X el cuadrado del dap. Se seleccionaron 20 árboles (de Lenga) y a cada uno se le midió el dap y el volumen. Los datos son:

$\Sigma x = 4,5862$ $\quad \Sigma y = 25,3520$ $\quad n = 20$
$\Sigma x^2 = 1,5487$ $\quad \Sigma y^2 = 46,4083$ $\quad \Sigma xy = 8,4464$

SCDX= 0,4970 SCDY= 14,2721 SPXY= 2,6329

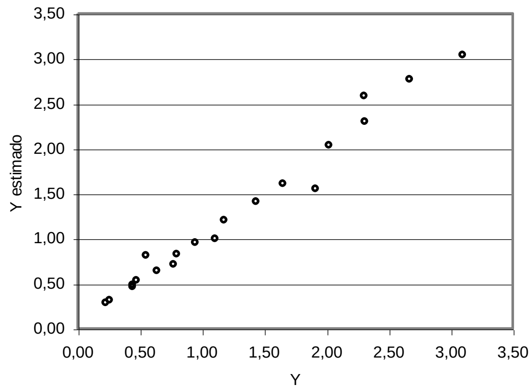
b = 2,6329/0,4970 = 5,2976

a = (1/20) [25,3520-(5,2976)(4,5862)] = 0,0528

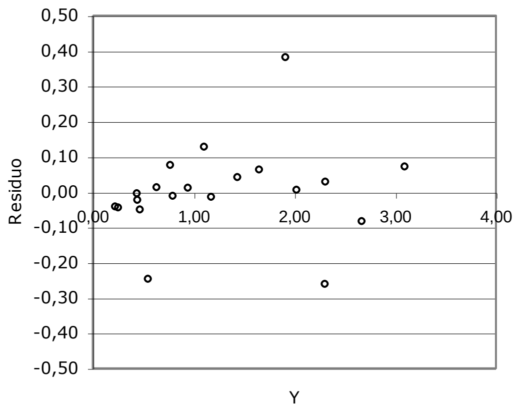
$$SCDReg = SCDY - \frac{SPXY^2}{SCDX} = 0,3241$$

$R^2$ = 1- (0,3241/14,2721) = 0.9773

El siguiente gráfico muestra el comportamiento de los valores estimados en función de los observados. Lo ideal es que los puntos se distribuyan siguiendo una línea recta con pendiente de 45 grados.




La siguiente figura muestra los residuos en función de las Y observadas.




Lo ideal es que se distribuyan en forma uniforme alrededor del cero.

## 2 TABLAS/ECUACIONES DE VOLUMEN

### 2.1 ANTECEDENTES

En el año 1821, el forestal alemán H. Cotta establece una serie de postulados básicos sobre el volumen de los árboles, que permanecen vigentes hasta hoy día. Esos postulados son:

a) que el volumen de un árbol es función de su diámetro (a la altura del pecho), de su altura y de su forma;

b) que si el volumen de un árbol ha sido correctamente establecido, ese volumen es también válido para cualquier otro árbol de igual diámetro, igual altura e igual forma; y

c) que mientras el diámetro y la altura son magnitudes medibles, la forma sólo puede ser comparada y juzgada.

Esto nos permite aceptar que si definimos una función o modelo que predice el volumen de madera para un árbol de un determinado dap, una determinada altura y una determinada forma, esa fórmula o modelo es válida para otro árbol del mismo dap, la misma altura y la misma forma.

No obstante, es posible recurrir a relaciones que no contengan indicadores de la forma; Clutter et al. dan las siguientes razones para preferir este tipo de relaciones:

a) la medición de diámetros superiores mediante trepada o dendrómetro, para establecer un cociente de forma, consumen mucho tiempo y dinero;

b) la variación en la forma de los troncos tiene mucho menos impacto sobre el volumen que la variación del dap o la altura;

c) en ciertas especies la forma es casi constante para cualquier tamaño de árbol; y

d) en otras especies, la forma del tronco está altamente correlacionada con el dap y la altura del árbol, por lo que el estas variables alcanzan a explicar gran parte de la variación del volumen debido a los cambios de forma.

Históricamente, el instrumento utilizado para expresar el volumen en función de una o más variables predictoras recibió el nombre genérico de Tablas de Volumen.

Una Tabla de Volumen es una presentación en forma de tabla que muestra el volumen medio o esperado por árbol, para árboles que comparten una o más variables predictoras. Las tablas se definen de la siguiente manera:

- Tabla local o de simple entrada: recurren sólo al dap, lo que implica asumir una relación constante dap/altura y dap/forma.
- Tabla estándar o de doble entrada: recurren al dap y a alguna altura, generalmente la comercial o total.
- Tabla de forma o de triple entrada: recurren al dap, a alguna altura y a algún indicador de la forma del tronco, generalmente un cociente de forma.

A lo largo del tiempo se han empleado tres métodos básicos para construir tablas de volumen, los que se indican en orden a su evolución:

a) Mediante volúmenes medios por clase

Se definían clases de diámetro y dentro de cada clase se seleccionaba una muestra de árboles a los que se les medía el volumen; finalmente se estimaba el volumen medio por árbol para cada clase. Así, todos los árboles pertenecientes a una misma clase diamétrica compartían el mismo volumen medio calculado.

b) Mediante el método gráfico

Fue posterior. Se procedía como en el caso anterior, pero los volúmenes medios se volcaban a un gráfico sobre el cual se trazaba, a mano, la correspondiente curva. Ya trazada la curva se leía en el eje Y el volumen medio estimado para cada clase diamétrica (X).

c) Mediante modelos de Regresión

Con los avances de la ciencia y la computación, la curva dibujada a mano en el método gráfico fue reemplazada por la curva expresada por un modelo matemático correspondiente a un Modelo de Regresión; aquí aparece la expresión *Ecuación de Volumen*. Los valores se pueden presentar en forma de ecuación o de tabla, por lo que cualquier ecuación de volumen recibe también el nombre de Tabla de Volumen. Más aún, el término Tabla de Volumen hoy se define como una función, tabla o gráfico que permite predecir el volumen de un árbol a partir de características como dap, altura y forma.

## 2.2 ECUACIONES DE VOLUMEN

### 2.2.1 Esquema general

La técnica hoy empleada para el desarrollo de ecuaciones de volumen es la de los Modelos de Regresión, consistente en definir una función matemática que permita predecir el volumen medio por árbol, para una especie o conjunto de especies, en función de una o más variables predictoras.

La fórmula matemática expresada en términos de variables abstractas, como $Y = f\ (Xi)$, recibe el nombre de *Modelo Matemático*, donde Y representa a la variable dependiente y Xi las variables independientes.

Cuando cada variable abstracta del modelo matemático recibe un significado concreto, decimos que operamos con el *Modelo Estadístico*, donde Y pasa a ser la variable respuesta y las Xi las variables predictoras que se presume son capaces de predecir razonablemente a la variable Y.

En nuestro caso, la variable respuesta es el volumen de interés del árbol y cada variable predictora es una variable medible en el árbol. Sin embargo, es común que el modelo contenga alguna función de estas variables, por ejemplo,

$$\log (Vol) = a + b \log (Dap)$$

Una vez resuelto el modelo podemos conocer el volumen esperado de un árbol si conocemos su dap, aun cuando la variable predicha es, en este caso, el logaritmo del volumen.

Finalmente, al modelo estadístico seleccionado para predecir el volumen medio recibe el nombre de Ecuación de Volumen.

### 2.2.2 Desarrollo de una ecuación de volumen

A continuación se describen brevemente los pasos básicos en el desarrollo de una ecuación de volumen.

i. Definición del volumen a predecir. El primer paso consiste en definir qué volumen habrá de predecir la ecuación. En general, hay dos posiciones extremas: a) definir el volumen asumiendo un destino específico, y b) definir un volumen general. Cada una tiene sus ventajas y desventajas.

ii. Definición de las variables de entrada; lo que implica seleccionar el tipo de tabla a construirse.

iii. Selección y medición de árboles muestra. Se debe definir cómo se seleccionarán los árboles muestra, qué mediciones se harán y con qué metodología. Con relación a la selección de los árboles muestra, las cuestiones básicas a tener en cuenta son tres:

- **Dónde** obtener la muestra. La muestra debe cubrir toda el área de interés, por lo que se deberá establecer un mecanismo de cobertura acorde con esta necesidad.

- **Cómo** obtener la muestra. Se debe obtener de manera que quede cubierto todo el rango de cada variable de entrada a usarse; sólo el muestreo de la variable respuesta es aleatorio.

- **Cuántos** árboles medir. No hay reglas fijas, pero a mayor número más precisa será la tabla; como referencia, se asume que para un rodal monoespecífico y homogéneo son suficientes entre 50 y 100 árboles para una tabla local y entre 80 y 150 para una tabla estándar.

iv. Ensayo de modelos matemáticos y selección del modelo. Si es posible, es recomendable comenzar con la representación gráfica de las variables observadas o de alguna función de ellas (para más de una variable independiente es más complicado lograr una representación gráfica). La forma de distribución de la nube de puntos nos da una idea del comportamiento de las variables probadas. Por ejemplo, si percibimos que la relación entre el cuadrado del dap ($d^2$) y el volumen (V) es lineal, un modelo del tipo $V = a + b.d^2$ es aceptable. En caso de dudas se pueden ensayar distintos modelos. En el caso particular de una tabla estándar, las variables dap (d) y altura (h) pueden operarse en forma independiente o en la forma que se conoce como variable combinada: $d^2h$. Se debe definir el método de estimación de parámetros.

v. Análisis de la calidad de la regresión. Estimados los parámetros de cada modelo en ensayo, debemos contar con algún indicador de su comportamiento, para lo cual recurrimos a una serie de indicadores estadísticas. Los indicadores numéricos más usados son el coeficiente de determinación $R^2$ y el análisis de los volúmenes estimados y de los residuos. La conversión posterior del modelo a tabla es opcional.

## 2.3 MODELOS DE POSIBLE USO EN EL DESARROLLO DE TABLAS DE VOLUMEN

- Para tabla de simple entrada:

$V = b_1 \; D^2$
$V = b_0 + b_1 \; D^2$
$V = b_0 + b_1 D + b_2 D^2$
$V = b_0 \; D^{b1}$

- Para tabla de doble entrada:

$V = b_1 \; D^2H$
$V = b_0 + b_1 \; D^2H$
$V = b_0 \; D^{b1} \; H^{b2}$

- Para tabla de triple entrada

$V = b_0 + b_1 \times D^2 \; H \; F$

## 2.4 CALCULO DEL VOLUMEN EN UN CONJUNTO DE ARBOLES MEDIANTE TABLAS DE VOL

Las ecuaciones y tablas de volumen se aplican para estimar el volumen de madera contenido en un conjunto de árboles, donde el volumen total es igual a la suma de los volúmenes de los árboles individuales.
Si disponemos de una ecuación de volumen, sus variables predictoras podrán medirse en cada árbol y estimarse para cada uno el correspondiente volumen; su suma dará el volumen de todos los árboles. Supongamos, por ejemplo que la ecuación de volumen a utilizar es la siguiente: Volumen $= a + b.\,D^2$ y que pretendemos determinar el volumen de madera contenido en un conjunto de **n** árboles; dicho volumen surge como:

$$\text{VolumenTotal} = \sum_1^n (a + b.D^2) = \sum_1^n a + \sum_1^n b.D^2 = n.a + b.\sum_1^n D^2$$

Un caso muy común de uso de tablas de volumen es en parcelas de muestreo, donde la tabla permite determinar el volumen del conjunto n de árboles presentes en cada parcela. Relacionando este volumen con el área de esa parcela podemos expresar el volumen en términos de volumen por hectárea: $m^3$/ha. Cuando la ecuación de volumen es de doble entrada la altura no suele medirse en todos los árboles involucrados, sino que se suele ajustar una función que permita predecir la altura en función del dap. Así, a cada árbol cuya altura no le fue medida se le asigna la proveniente de esa relación. El mismo procedimiento permite convertir una tabla de doble entrada en una de simple entrada.

## 2.5 CURVAS HIPSOMÉTRICAS

Como la medición de alturas de árboles es un procedimiento lento, es usual tratar de reducirlas a una mínima cantidad. Una forma para hacerlo son las curvas hipsométricas, que está representada por un modelo de regresión que expresa el valor esperado del árbol en función de su Dap. Algunos modelos disponibles son:

- $H = 1{,}30 + bD + cD^2$
- $H = 1{,}30 + H\,(1 - e^{-a\,D})$
- $\text{Log } H = a + b \text{ Log } D$
- $H = a + b \text{ Log } D$

Una vez determinada la función, se le asigna a cada árbol la altura que define la función. En el caso de un modelo de doble entrada, se reemplaza la altura por el modelo hipsométrico desarrollado.