

UNIVERSIDAD DE CONCEPCION  
FACULTAD DE CIENCIAS FORESTALES  
Departamento de Manejo de Bosques y Medio Ambiente



**SOFTWARE PARA LA VALIDACION DE SUPUESTOS DE  
REGRESION APLICADOS A FUNCIONES DE VOLUMEN Y AHUSAMIENTO**

Por

Nelson Enrique Agurto Pavez

MEMORIA DE TITULO PRESENTADA  
A LA FACULTAD DE CIENCIAS  
FORESTALES DE LA UNIVERSIDAD  
DE CONCEPCION PARA OPTAR AL  
TITULO DE INGENIERO FORESTAL.

CONCEPCION - CHILE  
1997

UNIVERSIDAD DE CONCEPCION  
FACULTAD DE CIENCIAS FORESTALES  
Departamento de Manejo de Bosques y Medio Ambiente



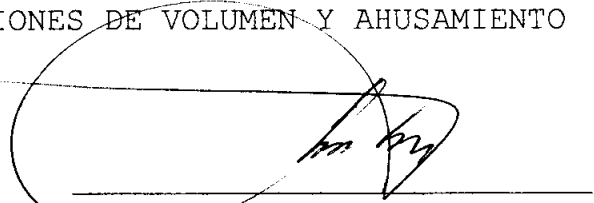
**SOFTWARE PARA LA VALIDACION DE SUPUESTOS DE  
REGRESION APLICADOS A FUNCIONES DE VOLUMEN Y AHUSAMIENTO**

MEMORIA DE TITULO PRESENTADA  
A LA FACULTAD DE CIENCIAS  
FORESTALES DE LA UNIVERSIDAD  
DE CONCEPCION PARA OPTAR AL  
TITULO DE INGENIERO FORESTAL.

CONCEPCION - CHILE  
1997

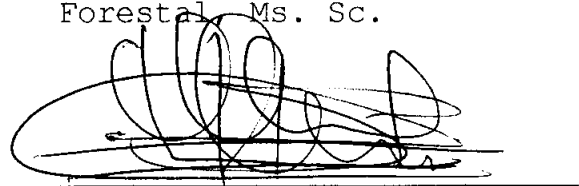
SOFTWARE PARA LA VALIDACION DE SUPUESTOS DE  
REGRESION APLICADOS A FUNCIONES DE VOLUMEN Y AHUSAMIENTO

Profesor Asesor.



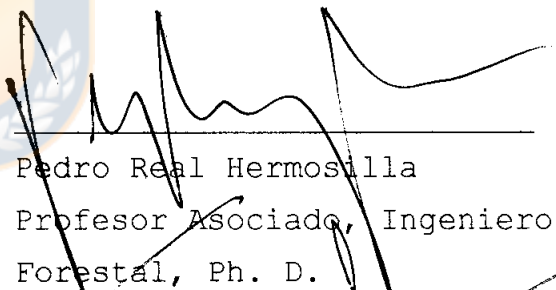
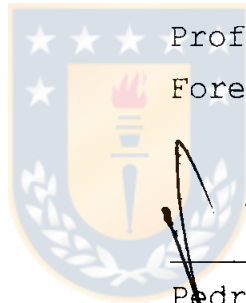
Jorge Cancino Cancino  
Profesor Asociado, Ingeniero  
Forestal, Ms. Sc.

Profesor Asesor.



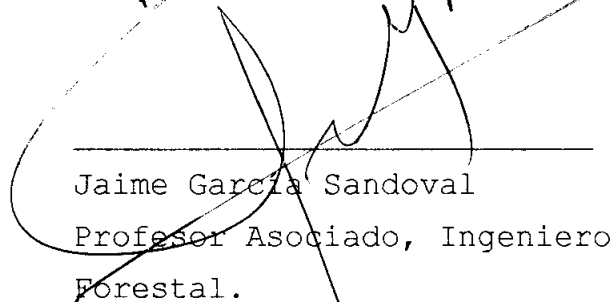
Manuel Lineros Parra  
Profesor Asociado, Ingeniero  
Forestal, Mg. Sc.

Director Dpto. Manejo de  
Bosques y Medio Ambiente.



Pedro Real Hermosilla  
Profesor Asociado, Ingeniero  
Forestal, Ph. D.

Decano Facultad de Ciencias  
Forestales.



Jaime Garcia Sandoval  
Profesor Asociado, Ingeniero  
Forestal.

Calificación de la memoria de título:

Sr. Jorge Cancino Cancino : Ochenta coma cero puntos.

Sr. Manuel Lineros Parra : Ochenta y nueve coma cuatro puntos.



A DIOS.

A MI ESPOSA Y BEBE.

A MI FAMILIA.

## INDICE DE MATERIAS

CAPITULOS	PAGINA
I	INTRODUCCION..... 1
1.1	Análisis de regresión y correlación..... 2
1.2	Los supuestos de regresión y sus implicancias estadísticas..... 5
1.3	Modelos de volumen..... 7
1.3.1	Funciones de volumen de árboles individuales.... 9
1.3.2	Funciones de volumen generales..... 10
1.4	Modelos de ahusamiento..... 14
1.4.1	Funciones de ahusamiento de árbol individual.... 16
1.4.2	Funciones de ahusamiento segmentadas..... 16
1.4.3	Funciones de ahusamiento generales..... 17
II	METODOLOGIA..... 18
2.1	Materiales..... 18
2.2	Programación..... 18
2.3	Estructura organizacional del sistema..... 18
2.4	Desarrollo e implementación del sistema..... 19
III	RESULTADOS..... 20
3.1	El sistema..... 20
3.1.1	Ingreso de datos..... 20
3.1.2	Modificación de los datos..... 22
3.1.3	Procesos..... 22
3.1.3.1	Elección del modelo..... 23
3.1.3.2	Ajuste del modelo..... 23
3.1.3.3	Validación de supuestos..... 23
3.1.3.4	Generar resumen..... 24
3.2	Gestiones generales..... 24
3.3	Instructivo de uso..... 25

CAPITULOS	PAGINA
IV	DISCUSION..... 26
V	CONCLUSIONES..... 27
VI	RESUMEN..... 28
	SUMMARY..... 29
VII	BIBLIOGRAFIA..... 30
VIII	APENDICES..... 33
IX	ANEXOS..... 44



INDICE DE TABLAS

TABLA N°	PAGINA
<u>En el apéndice</u>	
1 A Datos de entrada para el ejemplo .....	34
2 A Cuadro resultado para el ejemplo.....	38



## INDICE DE FIGURAS

FIGURA N° PAGINA

En el texto

1 Arquitectura del sistema..... 18

En el apéndice

1A Gráfico de dispersión de datos..... 36

2A Gráfico de dispersión de errores versus valor  
estimado..... 37





## INDICE DE APENDICE

APENDICE N°		PAGINA
1	Ejemplo de funcionamiento del sistema.....	34
2	Interfaz resultado para el supuesto de normalidad....	39
3	Interfaz resultado para el supuesto de homocedasticidad.....	40
4	Interfaz resultado para el supuesto de autocorrelación.....	41
5	Interfaz resultado para el supuesto de multicolinealidad.....	42
6	Interfaz resultado para el supuesto de especificación.....	43



## INDICE DE ANEXOS

ANEXO N°		PAGINA
1	Modelos de volumen más usados.....	45
2	Modelos de ahusamiento más usados.....	46
3	Los supuestos de regresión y la prueba estadística para validarlos.....	47



## I INTRODUCCION.

En el mundo de hoy, cada vez se tiende más a la complementación de las ciencias; así por ejemplo, la medicina se vale de los progresos de la energía nuclear para solucionar algunas patologías, creando nuevos fármacos y nuevas formas de diagnosticar enfermedades. Conocidos y muy populares son también los simuladores, que se valen de la computación para llevar a cabo alguna actividad riesgosa sin la necesidad de poner en peligro a la persona que la realiza; un ejemplo de esto son los simuladores de vuelo. Como éstos existen muchos otros casos de complementación, los cuales tienen por finalidad el ir mejorando y haciendo más eficientes las gestiones de las empresas o particulares. La ingeniería forestal y particularmente la mensura forestal no queda al margen de esta tendencia mundial.

La mensura basa sus principios en la estadística; es ésta la que día a día, junto con el avance científico y tecnológico se hace cada vez más imprescindible y es una de las principales herramientas utilizadas para reflejar el desarrollo de cualquier actividad.

En este documento se presenta un software para el ajuste y validación de los supuestos de regresión aplicados a funciones de volumen y ahusamiento. La validación de los supuestos de regresión normalmente no se realiza; esto puede deberse a que el ajuste lo efectúan profesionales experimentados, que se basan en el conocimiento empírico para determinar el tipo de modelo a usar, y el cual estaría cumpliendo los supuestos de regresión que la literatura señala. El software que se presenta

en este documento permite realizar la validación de esos supuestos en forma ágil, constituyéndose por ello en una herramienta útil para estos efectos.

Dada la gran cantidad de modelos posibles de ajustar, resulta muy difícil saber a priori cuál es el más conveniente para una serie de datos particulares. Es en este punto donde el sistema que se presenta es de gran utilidad; entre otros, permite ingresar un modelo cualquiera y posteriormente validar los supuestos de regresión. Si los resultados de la validación son buenos, el modelo ingresado es el apropiado para realizar estimaciones; por el contrario, si los resultados son deficientes, es necesario usar otro modelo de estimación.

En la revisión bibliográfica que se expone a continuación se analizan los siguientes temas:

- Análisis de regresión y correlación.
- Los supuestos de regresión y sus implicancias estadísticas.
- Modelos de volumen.
- Modelos de ahusamiento.

### 1.1. Análisis de regresión y correlación.

El análisis de regresión estudia la dependencia que existe entre una variable dependiente y una o más variables explicativas o independientes, con el objeto de estimar el valor promedio poblacional de la primera en base a los valores conocidos en muestras repetidas de las segundas (Gujarati, 1981).

La regresión es una relación matemática que explica la dependencia que existe entre dos o más variables. Esta dependencia puede ser de dos tipos (Peña, 1992):

- a) Funcional o determinística (exacta).
- b) Estadística o estocástica (probabilística).

En el primer caso, la relación existente es exacta; ello significa que las variables independientes determinan totalmente el valor de la variable dependiente que se desea conocer.

En el segundo caso, la relación ya no es exacta, sino que existe un error en la predicción; esto significa que las variables independientes no explican totalmente el comportamiento de la variable dependiente. El error más el aporte de las variables independientes explican el comportamiento total de la variable dependiente (respuesta). Una relación de este tipo se escribe de la siguiente forma:

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \dots + b_k X_{ki} + e_i$$

donde:

$Y_i$  = variable aleatoria dependiente (estocástica).

$X_i$  = variables independientes.

$b_i$  = coeficientes de regresión (constantes).

$e_i$  = error estocástico.

$i = 1, 2, 3, \dots, n$

y su forma matricial es:  $Y = X \hat{\beta} + e$

donde:

$Y$  = matriz columna de orden  $(n \times 1)$ .

$X$  = matriz de orden  $(n \times k)$ .

$\hat{\beta}$  = matriz columna de orden  $(k \times 1)$ .

$e$  = matriz columna de orden  $(n \times 1)$ .

El desvío  $e_i$ , que refleja el aporte de todas las variables excluidas del modelo, no es posible determinarlo, así el modelo se transforma en una relación probabilística (Peña, 1992).

Existen varios tipos de modelos:

- Lineal  $Y_i = b_0 + b_i X_i + e_i$
- Exponencial  $Y_i = e^{(b_0 + b_i X_i)} + e_i$
- Recíproco  $Y_i = 1 / (b_0 + b_i X_i) + e_i$

La elección del modelo adecuado es el primer paso en el análisis de regresión. Luego surge el problema de cómo estimar los parámetros poblacionales (coeficientes de regresión). Existen varias formas de realizar esta labor, que en estadística se denomina ajuste del modelo (Ostle, 1983). Los métodos de ajuste más conocidos son:

- Mínimos cuadrados ordinarios (M. C. O.)
- Mínimos cuadrados ponderados (M. C. P.)
- Mínimos cuadrados generalizados (M. C. G.)
- Máxima verosimilitud (M. V.)
- Ajuste robusto (A. R.)

El método de M. C. O. minimiza la sumatoria de los errores al cuadrado, o sea:

$$\sum_{i=1}^n e_i^2 = \text{mínimo.}$$

donde:  $e_i = Y_i - \hat{Y}_i$   
 $Y_i$  = valor real.  
 $\hat{Y}_i$  = valor estimado.

El método de M. C. O. es el más usado para el ajuste de funciones de volumen y ahusamiento de árboles individuales (Bruce and Schumacher, 1950; Furnival, 1961; Loetsch et al., 1973).

## 1.2. Los supuestos de regresión y sus implicancias estadísticas

Marshall et al. (1995) señalan que para emplear el método de M. C. O. en forma satisfactoria se deben cumplir cinco supuestos. Si ellos se cumplen los estimadores de los parámetros de la regresión tienen las propiedades deseables de normalidad, insesgamiento, consistencia y eficiencia.

### Supuesto N° 1.

Los errores poseen una distribución normal, su promedio es cero y la varianza es constante.

Simbólicamente:  $e \approx N(0, \sigma^2)$

Cuando el tamaño de la muestra es pequeño, existe mayor probabilidad de que este supuesto no se cumpla, pero a medida que el tamaño muestral crece (en forma asintótica) las pruebas de hipótesis y las estimaciones de los parámetros son más confiables.

### Supuesto N° 2.

La varianza condicional de los errores es constante u homocedástica.

En símbolos:

$$E(e_i^2) = \sigma^2 \quad i = 1, 2, \dots, n$$

Si no se cumple este supuesto existe heterocedasticidad de varianza. Luego las pruebas t y F producirán resultados

inexactos y los intervalos de confianza no serán válidos, dado que las varianzas de los parámetros serán más grandes que las verdaderas.

Supuesto N° 3.

No existe autocorrelación entre los errores.

En símbolos:

$$E(e_i, e_j) = 0 \quad i \neq j$$

Si no se cumple este supuesto, los estimadores de M. C. O. continúan siendo insesgados, consistentes, pero dejan de ser eficientes (de varianza mínima). Luego, las pruebas de hipótesis t y F pierden exactitud.

Supuesto N° 4.

No existe multicolinealidad entre las variables independientes.

Dos o más variables independientes no pueden presentar una colinealidad (relación lineal) perfecta o casi perfecta (Gujarati, 1981). Si no se cumple este supuesto las varianzas de los estimadores estarán aumentadas y en el límite (colinealidad perfecta) serán infinitas. Luego los intervalos de confianza para los parámetros serán más amplios y poco confiables.

Supuesto N° 5.

El modelo de regresión está correctamente especificado, es decir no existe sesgo (error) de especificación.

Existen varios factores que pueden llevar a un sesgo de especificación; el más relevante es excluir variables independientes que aporten mucho a la predicción. Luego, las consecuencias de violar este supuesto están en directa relación



con el error de especificación cometido. De este modo, si el error es excluir una variable relevante del modelo, la estimación de los parámetros es sesgada e inconsistente, la varianza del error se estima en forma incorrecta y se invalidan los procedimientos usuales de pruebas de hipótesis.

Gujarati (1981) y Marshall et al. (1995) señalan que al cumplirse estos supuestos se obtienen estimadores: Insesgados; lo que significa que la esperanza matemática de éstos, para la totalidad de las muestras de tamaño "n" de una población, es igual al valor poblacional, eficientes; para muestras grandes o pequeñas sus varianzas son mínimas, consistentes; a medida que aumenta el tamaño de la muestra, los estimadores convergen a su verdadero valor poblacional y normales; los estimadores de los coeficientes de regresión y los valores estimados de la variable dependiente están normalmente distribuidos.

Los modelos de regresión lineal, en general, son de uso extensivo en el área forestal, para el ajuste de funciones de volumen, ahusamiento y otras aplicaciones. Se ha evaluado también la calidad predictiva de modelos lineales y no lineales; los resultados muestran que los primeros son mejores que los últimos (Yonghe, 1987).

### 1.3. Modelos de volumen.

Cada propietario de un predio (rodal) forestal necesitará determinar en algún momento cuánta madera tiene. La expresión básica para la cuantificación de madera más usada en la actualidad es el "volumen de madera" (Cailliez, 1980).

En términos generales se distinguen dos tipos de volumen: Volumen sólido total; que es el volumen del árbol entero y el

volumen aserrable; que es el volumen del árbol hasta un diámetro límite o diámetro comercial.

Para determinar el volumen de un rodal se puede medir directamente cada árbol y sumar estos valores, labor impracticable cuando el rodal es de gran tamaño (Cailliez, 1980; Husch et al., 1982). Lo habitual es estimar el volumen mediante métodos estadísticos.

Así, el manejador forestal se ve enfrentado a uno de los problemas más antiguos de la mensura forestal, el que concierne a la estimación, por medio de la construcción de funciones (ecuaciones) de volumen para distintas especies y sitios (Cunia, 1964; Green y Strawderman, 1985).

Para construir una función de volumen se deben seguir tres pasos generales (Loetsch et al., 1973): Primero; selección de un número suficientemente grande de árboles representativos de la población (150 a 170 árboles es un número adecuado), segundo; medición de las variables independientes y dependiente, para la derivación del modelo (función) de volumen y finalmente prueba de los distintos modelos (funciones) por medio de pruebas estadísticas, bondad de ajuste, etc. y selección del modelo de volumen óptimo.

Existen distintos tipos de funciones de volumen (Cailliez, 1980): Funciones de árboles individuales; estas entregan el volumen de un árbol en función de variables predictoras de fácil medición, como el dap, funciones agregadas; que entregan el volumen total por unidad de superficie del rodal y funciones locales parametrizadas; que entregan el volumen de un árbol

como función de variables del propio árbol y variables relacionadas con el rodal.

Cailliez (1980) señala que las variables independientes en una función de volumen deben ser poco numerosas y fáciles de medir, estar fuertemente correlacionadas con el volumen y débilmente correlacionadas entre sí. En suma, cada modelo (función de volumen) debería seguir el principio de Occam o principio de la parsimonia, que enuncia que el modelo se debe mantener tan simple como sea posible (Gujarati, 1981) (un listado de funciones de volumen se entrega en el anexo N° 1).

1.3.1. Funciones de volumen de árboles individuales. Las funciones de volumen de árbol individual entregan el volumen de un árbol en función de variables predictoras de fácil medición, como el dap, la altura total o comercial, la clase de forma, etc (Cailliez, 1980). Las primeras funciones de volumen de árbol individual fueron construidas por Cotta en 1804, y publicadas en 1821, bajo el nombre de tablas de volumen estándar (Spurr, 1952).

Dentro de las funciones de volumen de árbol individual se pueden distinguir las funciones basadas en una sola variable independiente, generalmente el dap, llamadas corrientemente "funciones de volumen local" o tarifas (Spurr, 1952; Loetsch et al., 1973; Cailliez, 1980; Husch et al., 1982).

Simbólicamente:

$$V = f(\text{dap}).$$

Estas funciones de volumen son únicas para un rodal particular y sólo es posible su uso en el área donde se recolectaron los datos para construirla, es decir rodales con características

homogéneas, igual especie, edad, manejo, etc (Spurr, 1952; Loetsch et al., 1973; Husch et al., 1982; Da Silva et al., 1992). La última afirmación se fundamenta en el hecho de que la razón altura-diámetro permanece uniforme en el área muestreada, es decir, no existen alturas significativamente diferentes para cada clase diamétrica (Spurr, 1952; Loetsch et al., 1973; Husch et al., 1982).

Una vez obtenida la función de volumen local, se puede determinar el volumen por hectárea y total del rodal ingresando el dap del árbol de área basal media en la función y luego multiplicar el resultado por el número de árboles respectivo. Este método da resultados sesgados porque los coeficientes de la ecuación no han sido influenciados por cada clase diamétrica. El método tenderá a serias subestimaciones del volumen, si la ecuación es para predecir volumen hasta un diámetro de corte y el diámetro promedio está cercano a ese límite (Cailliez, 1980).

1.3.2. Funciones de volumen generales. Aquellas funciones que presentan dos o más dimensiones del árbol como variables independientes son llamadas "funciones generales de volumen" (Spurr, 1952; Loetsch et al., 1973; Husch et al., 1982).

Simbólicamente:

$$V = f (dap, h, g, \dots)$$

Los regresores más usados en este tipo de funciones son el dap y la altura total (h), variables que participan en los modelos de regresión a la forma de diferentes transformaciones. Algunos modelos incorporan además la forma del árbol, mediante el uso de algún factor, pero Behre y Smith (1935), Ker y Csizmazia

(1961), concluyen que no existe ganancia práctica al introducir la forma del árbol.

Trabajos realizados por Pollanschütz (1965) (citado por Max y Burkhart, 1976) muestran que al introducir al modelo una tercera variable ( $D_i$ : diámetro a una altura  $i$ ) y al compararlo con modelos de volumen que sólo usan dos variables independientes ( $d$ ,  $h$ ), aquélla produce un descenso del valor de la desviación estándar del volumen en más de un 50%, en la mayoría de los casos.

Este tipo de funciones es de uso más general que las tarifas. Pueden abarcar áreas más extensas, pero se debe tener cuidado al utilizarlas en regiones alejadas de la zona en donde se confeccionó; factores como sitio, manejo y otros pueden alterar severamente su utilidad. Luego, resulta imposible recomendar un modelo único, para el caso de funciones de volumen (Cailliez, 1980); sin embargo, la existencia de modernos y sofisticados sistemas de procesamiento electrónico de datos, a la par de numerosos programas computacionales para la selección de modelos y ajustes de funciones, convierten la tarea de encontrar la mejor función de volumen en una labor más sencilla que antaño.

Una de las principales dificultades que presenta el ajuste de funciones de volumen, cualquiera sea su tipo, es el problema de la varianza. Se ha podido constatar que la varianza del volumen varía considerablemente con el tamaño de los árboles (Furnival, 1961; Cunia, 1964); esto implica recomendar el uso sistemático de regresión ponderada (Cailliez, 1980). La ponderación busca homogenizar la varianza del volumen. Ello se logra mediante la determinación de una función que represente la relación entre

la varianza y las variables independientes que participan en el modelo, empleando luego el recíproco de la función generada como factor de ponderación (Husch et al., 1982; Green y Strawderman, 1985; Parresol, 1993).

Si el supuesto de homocedasticidad no se cumple, es decir la varianza del volumen se comporta en forma heterocedástica, y no se pondera, el error de estimación resulta inconsistente e inconstante y las pruebas de hipótesis que se realicen generarán resultados erróneos.

Una solución, como ya se mencionó, es ponderar la varianza del volumen por algún factor que minimice el efecto de la heterocedasticidad, pero en la práctica es muy difícil determinar una ponderación apropiada para una función particular (Furnival, 1961). Otra forma de solución es realizar alguna transformación a la variable dependiente; estas transformaciones pueden ser de tipo logarítmica, raíces u otras (Cunia, 1964; Gujarati, 1981; Husch et al., 1982). Sin embargo, las funciones de volumen del tipo logarítmico, donde la variable dependiente es "log V" en vez de "V", subestiman sistemáticamente el volumen, pues el siguiente modelo:

$$\log V = \log b_0 + b_1 \log (D)$$

derivado de:

$$V = b_0 D^{b_1}$$

que predice a "log V" y no a "V", entrega una estimación del promedio geométrico (y no del aritmético) de los volúmenes de los árboles de diámetro "D" (Cunia, 1964).

Gujarati (1981) señala que el problema de heterocedasticidad es más frecuente en series de corte transversal que en series de tiempo. Esta afirmación es relevante pues en la construcción de funciones de volumen se trabaja con este tipo de información, es decir, datos tomados en un momento determinado del tiempo. Señala, además, que los estimadores obtenidos bajo estas circunstancias siguen siendo lineales, insesgados y consistentes, pero dejan de ser eficientes (no poseen varianzas mínimas), ni siquiera asintóticamente (con muestras grandes) (Gregoire y Dyer, 1989; Parresol, 1993; Marshall et al., 1995).

Gujarati (1981), Carroll y Spiegelman (1992) y Marshall et al. (1995) mencionan que no existen reglas fijas para detectar la heterocedasticidad, pero indican que la mayoría de los métodos destinados para ello se basan en el examen de los errores ( $e_i$ ) de M.C.O., (esto si los errores de la muestra son buenas estimaciones de los errores poblacionales, lo que se puede cumplir si el tamaño de la muestra es relativamente grande). Al respecto, Gujarati (1981) señala que pruebas como las de Park, Glejser, Goldfeld y Quandt pueden ser usadas como métodos informales y no deben ser consideradas como concluyentes al respecto, pues cada una de ellas se fundamenta en una serie de supuestos que dependerán de la naturaleza particular de los datos con los que se trabaja y de la severidad de la heterocedasticidad.

LeMay (1990) menciona que los términos de error " $e_i$ " en funciones simples (no simultáneas) son independientes y están idénticamente distribuidos y poseen varianza constante, pero en las funciones de volumen este supuesto no se cumple; esto resulta en que las varianzas estimadas de los coeficientes son estadísticamente sesgadas y no consistentes y los intervalos de

confianza para los coeficientes y los valores medios estimados no son exactos. Así mismo, propone la técnica de M.C.O. en varias etapas o MSLS (multistage least squares) para solucionar el problema antes señalado, obteniendo estimadores consistentes de los coeficientes y de sus varianzas asociadas.

#### 1.4. Modelos de ahusamiento.

Un modelo de ahusamiento o modelo fustal es una relación matemática que expresa el diámetro con o sin corteza a cualquier altura del fuste en función de variables de estado de fácil medición en el árbol (Allen, 1992).

Aunque la forma del árbol no puede ser descrita completamente mediante una función matemática, es común y conveniente asumir que la forma del árbol se aproxima a varios sólidos geométricos (Liu, 1980; Husch et al., 1982). La porción inferior del fuste tiene la forma de un tronco de niloide, la porción intermedia la de un tronco de paraboloides, y la superior la de un cono (Max y Burkhart, 1976). Sin embargo, Grosenbaugh (1966, citado por Liu, 1980) afirma que cualquier forma funcional preconcebida no representa adecuadamente el ahusamiento del fuste, basándose en la observación de árboles individuales. Estos son capaces de asumir una variedad infinitas de formas, por lo cual una definición analítica y explícita de la forma del árbol carece de generalidad. El fundamento de esta afirmación, según Liu (1980), es que el ahusamiento o forma del fuste es un factor inestable y sensitivo a muchas interacciones en un sistema forestal dinámico. Existe evidencia biológica, analítica y visual de que árboles de la misma especie, que crecen en el mismo sitio y que reciben tratamientos similares, asumen una forma distinta.



No obstante esto, los variados objetivos del manejo forestal exigen la estimación de volúmenes de productos rollizos de diversas dimensiones; así, los manejadores forestales tratan de modelar mediante alguna relación matemática el perfil del árbol, con el objetivo de estimar con precisión volúmenes del fuste hasta ciertos diámetros límites o comerciales. De ahí la gran importancia que revisten los modelos fustales; gracias a su flexibilidad, y mediante el uso de herramientas de cálculo diferencial e integral, permiten estimar diámetros a diferentes alturas y volumen entre límites comerciales variables.

Los primeros estudios de la forma del fuste indicaron que los árboles seguían una forma geométrica conocida. Höjer (1903, citado por Demaerschalk, 1972) fue quien primero representó la forma del fuste mediante un modelo matemático. Luego aparecieron más y mejores modelos que se acercaban más al verdadero perfil del árbol.

Demaerschalk (1972) hizo un notable avance en la teoría de las curvas de ahusamiento al desarrollar las llamadas "funciones de ahusamiento compatibles". El término compatible se debe a que la integración de la función de ahusamiento genera un valor total igual al que entrega la función de volumen de la cual fue derivada (sin embargo no es la única forma de desarrollar una función de ahusamiento). Luego, la exactitud y precisión de la función de ahusamiento depende en forma directa de la exactitud y precisión de la función de volumen que se utilizó para construirla. Entre las funciones de ahusamiento se pueden distinguir tres grupos, ellos son:

- Árboles individuales.
- Segmentados.
- Generales.

1.4.1. Funciones de ahusamiento de árbol individual. Estas funciones son construidas ajustando una función para cada árbol de la muestra, para luego desarrollar funciones estimadoras de sus coeficientes. De esta manera el modelo fustal de cada árbol queda determinado ajustando sus coeficientes con las funciones estimadoras de coeficientes. Cailliez (1980) utiliza un polinomio de tercer grado para representar el perfil de árboles individuales y expresa la curva de perfil como:

$$S = f ( H )$$

donde:

S = área de la sección del fuste a la altura H.

H = altura relativa del árbol.

Señala que de esta forma es más fácil la obtención de volúmenes a cualquier altura que usando una relación del tipo:

$$D = f ( H )$$

donde:

D = diámetro del fuste a la altura H.

H = altura relativa del árbol.

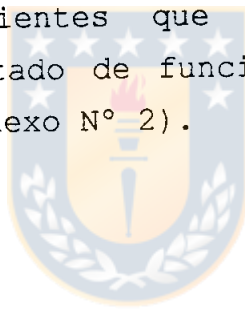
donde los cálculos de volumen son más largos y menos precisos debido a la transformación de unidades.

1.4.2. Funciones de ahusamiento segmentadas. Estas funciones representan la forma del fuste por medio de dos o más ecuaciones las que son forzadas a ser continuas en los puntos de unión de las mismas. Max y Burkhart (1976) señalan que la técnica de regresión segmentada aproxima el perfil del fuste de mejor manera que los modelos cuadráticos que no son segmentados, utilizados extensamente por otros autores.

1.4.3. Funciones de ahusamiento generales. Estas funciones son las más comunes; entre ellas encontramos las funciones polinomiales que se construyen mediante diferentes métodos, con diferentes variables independientes y dependientes, y exponentes que pueden llegar a sobrepasar el grado 50, llamadas también "funciones potencia".

Los modelos de ahusamiento generales representan el perfil promedio del fuste para toda una población (Forslund, 1990).

Los últimos estudios señalan que los modelos sigmoidales complejos son más eficientes que las funciones potencia (Forslund, 1990) (un listado de funciones de ahusamiento más usadas se entrega en el anexo N° 2).



## II METODOLOGIA.

### 2.1. Materiales.

En el desarrollo del sistema se utilizó el siguiente hardware: un computador con microprocesador 486 DX4 de 100 Mhz; un disco duro de 850 Mb; 4 Mb de memoria RAM y un ratón. Implementado con: sistema operativo MS-DOS 6.22; administrador de programas Windows 3.11 y lenguaje de programación Visual Basic 3.0.

### 2.2. Programación.

La programación se hizo en lenguaje computacional VISUAL BASIC 3.0. Este es uno de los primeros lenguajes que admite la programación orientada a objetos y eventos, que es un estilo de programación especialmente adaptado a interfaces gráficas del usuario y corre bajo ambiente Windows.

### 2.3. Estructura organizacional del sistema.

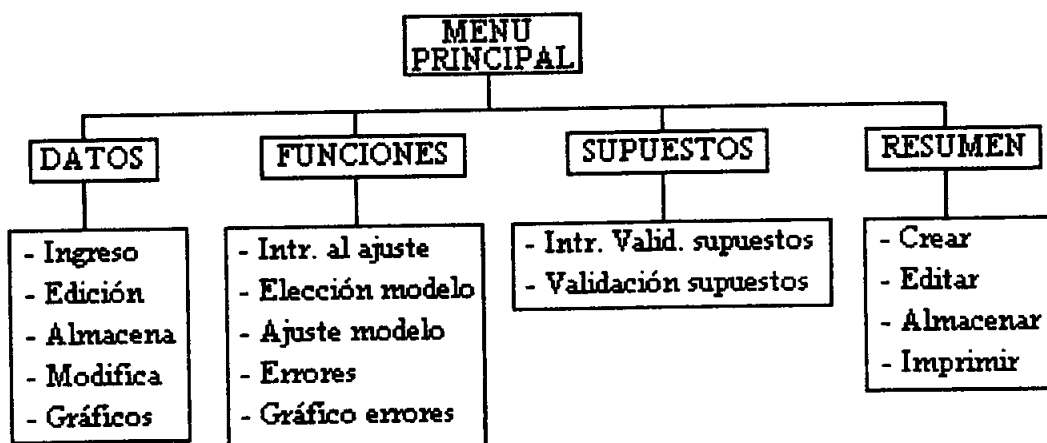


FIGURA N° 1. Arquitectura del sistema.

Esta estructura organizacional representa la jerarquía y orden de los diferentes módulos y de las distintas rutinas que conforman el sistema. La estructura del sistema consta de tres niveles: el primer nivel corresponde al menú principal; donde se elige uno de los módulos dependiendo de la rutina que se desee desarrollar; el segundo nivel presenta las diferentes rutinas del sistema; el tercer nivel ejecuta la rutina elegida.

Cada módulo y rutina posee un ítem de ayuda al instante que señala el proceso en el cual se está trabajando o se va a trabajar. Además, la rutina de validación de supuestos trae incorporada una ayuda más completa, para informar de forma general al usuario sobre los conceptos con los cuales está trabajando.

#### 2.4. Desarrollo e implementación del sistema.

En el desarrollo e implementación del sistema se trabajó en forma conjunta con un profesional de Ingeniería Civil en Informática de la Universidad de Concepción. Se conformó entonces un grupo interdisciplinario en materias de informática y de ciencias forestales que se abocó a la realización del sistema. Para esta labor se ocupó la metodología sugerida por Booch (1996), esto debido a que en la etapa de análisis de factibilidad se tuvo como requisito principal del sistema el disponer de una interfaz gráfica del tipo Windows y una programación orientada a objetos, ya que ésta presenta un menor grado de dificultad para el usuario final.

### III RESULTADOS.

El software para el ajuste y validación de supuestos de regresión es un sistema nuevo y orientado a necesidades específicas del área forestal. Este opera en un computador personal (piénsese en un 386 DX2 de 40 Mhz , 50 Mb en disco duro y un monitor VGA como mínimo) y no está sujeto a configuraciones especiales ni a la presencia de otros sistemas

#### 3.1. El sistema.

3.1.1. Ingreso de datos. Este punto considera tres aspectos:

a) Identificación del predio. Se solicita al usuario que ingrese los datos del predio, del propietario, y de la muestra. Estos datos son:

- Datos del predio; nombre del predio, rol S.I.I., comuna, provincia, región, superficie, especie arbórea, sitio, edad y origen de masa boscosa.
- Datos del propietario; nombre, dirección, fono casa y/o oficina.
- Datos de la muestra; número de árboles muestreados (se considera sólo aquellos árboles de un mismo predio).

b) Ingreso de datos. El sistema permite el ingreso de datos en la forma en que estos son obtenidos en terreno. El número válido de árboles muestreados oscilará entre los 20 y 400 (Loetsch et al. (1973) señalan que una muestra de 150 a 170 árboles es suficiente para realizar este tipo de estudios).

Los datos son obtenidos de acuerdo a dos modalidades: la primera es con mediciones a intervalos porcentuales de la longitud del fuste, y la segunda modalidad es a intervalos de

altura constante (cada cierto número de centímetros establecidos previamente). No se considera en esta versión del sistema el ingreso de datos para intervalos de amplitud variable.

Al sistema se ingresan los siguientes valores:

- Distancia de trozado, según:

i) altura porcentual, el usuario ingresa un porcentaje entre 5 y 50%.

ii) altura constante, el usuario ingresa una distancia de corte constante entre 1 y 5 m.

- Altura total del árbol.

- radio 1.

- radio 2.

- radio del dap.

c) Pre-procesamiento de datos. Una vez ingresado los datos, el sistema los procesa y arroja como resultado dos tipos de datos:

i) Datos de volumen; son aquellos que presentan un valor distinto para cada árbol, estos son:

- Volumen total (V); se calcula como sumatoria de los volúmenes de cada una de las secciones del árbol, mediante el uso de la fórmula de Smalian.

- Dap (D); se obtiene del promedio de los radios ingresados a esa altura del árbol.

- Altura total (H); esta medida siempre es obtenida en terreno.

- Clase de forma de Girard (F); depende de la clase diamétrica.

- Area basal (g); área de la sección perpendicular al eje del árbol a la altura del dap (1.3 m).

ii) Datos de ahusamiento; son datos que presentan más de una ocurrencia por árbol. El número de ocurrencias está determinado por la distancia de corte o trozado del árbol. Los datos son:

- Diámetro relativo; obtenido a partir de los radios ingresados.
- Altura relativa; depende de la longitud de las secciones.

3.1.2. Modificación de los datos. El sistema provee una interfaz que permite al usuario efectuar modificaciones sobre los datos que ya han sido ingresados. Estas modificaciones son del tipo:

- Agregar variable; puede efectuar alguna transformación sobre las variables pre-procesadas por el sistema. Las transformaciones son del tipo:

$$\text{Ln} ( X^a \text{ op } Y^b \text{ op } Z^c )$$

donde:

X, Y, Z son variables (v.g. volumen , dap, etc).

a, b,c = (0, 1, 2, 3, 1/2, 1/3).

op = {\*, / } (operadores aritméticos).

- Eliminar variable; sólo puede eliminar aquellas variables que el mismo usuario ha ingresado como transformación y no aquellas que resulten del pre-procesamiento de datos.

- Modificar valores; puede modificar los valores originales ingresados al sistema de aquellas variables pre-procesadas y no transformadas.

3.1.3. Procesos. Este punto se refiere a aspectos que tienen relación con el procesamiento de los datos, es decir, ajuste, validación de supuestos y salida de los datos previamente ingresados al sistema.



3.1.3.1. Elección del modelo. El usuario decide qué tipo de modelo (función) desea ajustar con los datos ingresados. Puede elegir entre funciones de volumen y de ahusamiento. En ambos casos puede ingresar la función que él desee, considerando las variables obtenidas del pre-procesamiento de datos, más las posibles transformaciones realizadas con la opción agregar variable (punto 4.2.2).

3.1.3.2. Ajuste del modelo. Después de ingresar la función (volumen o ahusamiento), el sistema la ajusta por el método de los mínimos cuadrados ordinarios y luego despliega los resultados del ajuste. Despliega también las estimaciones y errores asociados a cada valor real del árbol, obtenido del pre-procesamiento.

3.1.3.3. Validación de supuestos. El sistema valida los supuestos de regresión lineal y presenta al usuario los resultados de esta evaluación (El detalle de cada uno de estos supuestos y la prueba estadística usada para validarlos se puede encontrar en el anexo N° 3).

a) Validación; el sistema presenta al usuario los supuestos estadísticos, para que éste seleccione el que desea validar. Este proceso puede realizarse en forma reiterativa. El usuario además debe indicar el nivel de confianza con que desea trabajar (opciones: 90, 95, 99%).

b) Despliegue de resultados de la validación; el sistema despliega los resultados de la validación de cada uno de los supuestos, señalando el nombre del supuesto, su enunciado, el estadístico usado y su valor, el valor crítico y el resultado (cumple / no cumple).

c) Información sobre cada supuesto; el sistema da al usuario la opción de obtener información adicional sobre cada uno de los supuestos (v.g. implicancias del supuesto, medidas remediales, bibliografía de referencia).

3.1.3.4. Generar resumen. El sistema puede generar un resumen que contiene la información más relevante del ajuste y validación realizada. Además es capaz de almacenar e imprimir dicho resumen.

### 3.2. Gestiones generales.

En este punto se tocan dos aspectos que afectan a todo el sistema y que son considerados de importancia:

- Interfaz; la calidad y claridad de la interfaz es relevante. Se debe recordar que una de las razones que motiva la construcción del software es precisamente la carencia de un sistema estadístico amistoso para el usuario. Dada la importancia que reviste la interacción usuario-sistema, se consideró la interfaz gráfica del tipo Windows como la más adecuada.

- Validación de datos; para disminuir al máximo los posibles errores generados al ingresar o seleccionar datos, se contemplaron las siguientes medidas:

- i) Acotar valores; para que sólo sean aceptados como válidos los valores pertenecientes a un rango predefinido.
- ii) Ingresar seleccionando; se presenta una serie de valores o expresiones de donde sólo se deberá elegir la deseada.
- iii) Reingreso de datos; el sistema no permite seguir avanzando si un dato ha sido mal ingresado u omitido.

### 3.3. Instructivo de uso.

Para facilitar el acceso al sistema, en el apéndice N° 1 se muestra un ejemplo para el ingreso, operación y resultados que entrega el software.



#### IV DISCUSION.

En la actualidad existen diversos sistemas (software) que permiten construir una gran gama de funciones, utilizando distintos métodos de ajuste, pero existen muy pocos sistemas que realicen la validación de los supuestos de regresión lineal que deben cumplir éstas funciones.

La mayoría de estos sistemas requieren un buen conocimiento de los conceptos estadísticos por parte del usuario y si a eso le agregamos el idioma en que están implementados, se convierten en sistemas restrictivos para el uso generalizado de ellos.

El sistema aquí presentado es de uso específico para el sector forestal y su interfaz usuario-sistema lo hacen muy fácil de usar, junto con poseer rutinas de ayuda que facilitan aún más su operación y comprensión de resultados.

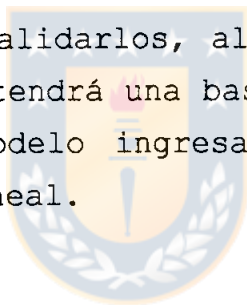
Las rutinas de validación de supuestos poseen una sola prueba estadística para validar cada supuesto, lo que puede hacer pensar que el sistema es débil en éste aspecto, sin embargo, el software se transforma en una herramienta de apoyo eficaz para el sector, si lo comparamos con la complejidad que implicaría la estructuración de las pruebas asociadas a los supuestos de regresión en los paquetes estadísticos convencionales disponibles hoy en el mercado.

## V CONCLUSIONES.

1.- El sistema sirve como herramienta de trabajo para la construcción de funciones de volumen y ahusamiento.

2.- El sistema permitiría un menor gasto de tiempo en desarrollar las actividades anteriores, lo que significaría una mejor gestión.

3.- Los resultados de la validación de supuestos no son concluyentes, esto debido al hecho de que sólo se utiliza una prueba estadística para validarlos, al existir más pruebas que validen cada supuesto se tendrá una base estadística más sólida para determinar si el modelo ingresado cumple o no con los supuestos de regresión lineal.



## VI RESUMEN.

El trabajo realizado en esta memoria de título fue la elaboración de un software en Visual Basic 3.0, el software ajusta y valida supuestos de regresión aplicados a funciones de volumen y ahusamiento.

Después de ingresar los datos de los árboles, el usuario ajusta un modelo de volumen o ahusamiento a su elección, por medio del método de mínimos cuadrados ordinarios. Luego puede efectuar la validación de los supuestos de regresión para determinar si el modelo ingresado cumple o no con éstos supuestos.

El sistema consta de varias facilidades para el usuario, entre las que se cuentan gráficos de dispersión, transformación de variables, eliminación de variables, cambio de valores originales de los datos, almacenamiento y edición de datos y de resultados, impresión de resultados. Posee un módulo de ayuda que facilita la comprensión y manejo del sistema, transformándose en una herramienta de apoyo para el aprendizaje en disciplinas como mensura, estadística y otras.

## SUMMARY

The work carried out in this memory title was the elaboration of a software in Visual Basic 3.0; the software adjusts and validates assumptions of regression applied to volume and taper functions.

After entering the data about the trees, the user adjusts a model of volume or taper to his selection, by means of the ordinary least squares method. After that he can test of the assumptions of regression in order to determine if the model entered complies or not with these assumptions.

The system has various facilities for the user, of which you can count on, such as dispersion graphics, variable transformation, elimination of variables, change in original values of the data, storage and edition of data and results, printing of results. It has a help module which facilitates the understanding and management of the system, transforming it into a tool of assistance for the learning in disciplines such as mensuration, statistics and another.

## VII BIBLIOGRAFIA.

- 1.- Allen, P.J., 1992. Polynomial taper equation for *Pinus caribaea*. N. Z. J. For. Sci. 21(2):194-205.
- 2.- Booch, G., 1996. Análisis y diseño orientado a objetos con aplicaciones. 2ª edición. Mass. 589 p.
- 3.- Bruce, D. and F. Schumacher, 1950. Medición forestal. Editorial Herrero. México. 475 p.
- 4.- Cailliez, F., 1980. Estimación del volumen forestal y predicción del rendimiento con referencia especial a los trópicos. Volumen I y II. F.A.O. Roma. 90 p.
- 5.- Carroll, R.J. and C.H. Spiegelman, 1992. Diagnostic for nonlinearity and heteroscedasticity in error in variables regression. Technometrics. 34(2):186-190.
- 6.- Cunia, T., 1964. Weighted least squares method and construction of volume tables. For. Sci. 10(2):180-187.
- 7.- Da Silva, J.A., B.E. Borders and G.H. Brister, 1992. A tree volumen equation based on two lower stem diameters for *caribaea* pine in Sri Lanka. Comm. For. Rev. 7(2):114-118.
- 8.- Demaerschalk, J.P., 1972. Converting volume equations to compatible taper equations. For. Sci. 18:241-245.



- 9.- Forslund, R.R., 1990. The power function as a simple stem profile examination tool. *Can. J. For. Res.* 21:193-198.
- 10.- Furnival, G.M., 1961. An index for comparing equations used in constructing volume tables. *For. Sci.* 7(4):337-341.
- 11.- Green, E.J. and W.E. Strawderman, 1985. The use of bayes/empirical bayes estimation in individual tree volume equation development. *For. Sci.* 31(4):975-990.
- 12.- Gregoire, T.G. and M.E. Dyer, 1989. Model fitting under patterned heterogeneity of variance. *For. Sci.* 35(1):105-125.
- 13.- Gujarati, D., 1981. *Econometría básica*. McGraw-Hill Latinoamericana S.A. Bogotá Colombia. 465 p.
- 14.- Husch, B., C.I. Miller and T.W. Beers, 1982. *Forest mensuration*. John Wiley and Sons. 3ª edición. 402 p.
- 15.- LeMay, V.M., 1990. MSLS; a linear least squares technique for fitting a simultaneous system of equations with a generalized error structure. *Can. J. For. Res.* 20:1830-1839.
- 16.- Liu, C.J., 1980. Log volume estimation with spline approximation. *For. Sci.* 26(3):361-369.
- 17.- Loetsch, F., F. Zöhrer and K.E. Haller, 1973. *Forest inventory*. Munchen Bern Wien 470 p.

- 18.- Marshall, P., T. Szikszai, V. LeMay and A. Kozak. 1995. Testing the distributional assumptions of least squares linear regression. *For. Chron.* 71(2):213-217.
- 19.- Max, T.A. and H.E. Burkhart, 1976. Segmented polynomial regression applied to taper equations. *For. Sci.* 22(3):283-289.
- 20.- Ostle, B., 1983. *Estadística aplicada*. Editorial Limusa. México. 630 p.
- 21.- Parresol, B.R., 1993. Modeling multiplicative error variance; an example predicting tree diameter from stump dimensions in baldcypress. *For. Sci.* 39(4):670-679.
- 22.- Peña, D., 1992. *Estadística modelos y métodos*. Vol. II. Modelos lineales y series temporales. Editorial Alianza. 742 p.
- 23.- Seber, G., 1977. *Linear regression analysis*. John Wiley and Sons. N. York. 465 p.
- 24.- Spurr, S.H., 1952. *Forest inventory*. The Roland Press Company. N. York. 476 p.
- 25.- Yonghe, W., 1987. Tree volume equations; a statistical examination (summary). *For. Abs.* 48(4):249.



VIII APENDICES.

APENDICE N°1: Ejemplo de funcionamiento del sistema.

A continuación se presenta un conjunto de datos que servirán para mostrar las diferentes rutinas y los resultados que entrega el sistema. Los radios son medidos a la altura de tocón (20 cm), del Dap (sin corteza)(1.3 m) y a intervalos de distancia de 3 m.

TABLA 1A: DATOS DE ENTRADA PARA EL EJEMPLO.

ARBOL N°	A. TOTAL	DISTANCIA DE CORTE (m).							
		TOCON	1.3	3	6	9	12	15	18
01	18.1	12.8	10.5	9.1	7.5	5.2	3.5	1.1	0.4
		12.5	10.2	9.0	7.6	5.3	3.7	0.9	0.5
02	16.7	11.8	10.8	9.0	7.8	5.4	3.2	1.2	
		11.9	10.5	9.3	7.9	5.6	3.5	1.3	
03	13.1	10.3	9.5	7.2	5.1	3.6	1.0		
		10.1	9.2	7.5	5.2	3.4	0.9		
04	15.3	12.4	10.1	9.1	7.6	5.5	3.3	1.1	
		12.6	10.2	9.1	7.7	5.4	3.1	1.4	
05	19.1	13.9	11.2	9.5	7.1	5.4	3.1	1.0	0.5
		13.6	11.5	9.4	7.2	5.1	3.4	0.9	0.5
06	10.2	10.1	7.5	5.5	3.5	1.0			
		10.2	7.2	5.5	3.4	1.2			
07	20.4	15.4	13.1	10.4	7.3	5.4	3.7	1.2	0.5
		15.0	13.2	10.0	7.6	5.3	3.7	0.9	0.4
08	11.7	9.5	7.1	5.5	3.1	1.1			
		9.2	7.3	5.4	3.3	1.2			
09	13.6	10.6	9.1	7.7	5.5	3.1	1.3		
		10.2	9.2	7.7	5.5	3.4	1.1		
10	18.3	12.4	10.1	9.3	7.4	5.1	3.4	1.0	0.5
		12.9	10.1	9.2	7.3	5.2	3.3	1.0	0.4

continuación tabla 1A.

		DISTANCIA DE CORTE (m).							
ARBOL N°	A. TOTAL	TOCON	1.3	3	6	9	12	15	18
11	15.3	12.8	11.5	9.2	7.3	5.1	3.2	1.0	
		12.5	11.2	9.1	7.5	5.0	3.1	0.9	
12	19.2	14.5	12.6	9.1	7.5	5.5	3.1	1.1	0.4
		14.4	12.3	9.4	7.6	5.7	3.0	1.0	0.4
13	11.4	9.2	7.1	5.1	3.2	1.6			
		9.3	7.2	5.0	3.4	1.7			
14	14.3	10.9	9.1	7.5	5.1	3.3	1.1		
		10.4	9.4	7.2	5.4	3.3	1.3		
15	16.9	11.8	10.3	9.0	7.2	5.0	3.0	1.0	
		11.7	10.2	9.0	7.3	5.1	3.1	0.9	
16	10.5	9.1	7.5	5.2	3.1	1.1			
		9.3	7.3	5.2	3.2	1.2			
17	14.3	10.2	9.3	7.4	5.0	3.3	1.0		
		10.2	9.2	7.1	5.1	3.2	0.9		
18	15.5	12.8	11.8	9.2	7.7	5.3	3.0	1.0	
		12.9	11.5	9.1	7.7	5.4	3.1	1.1	
19	18.4	14.6	11.6	9.3	7.3	5.2	3.1	1.0	0.4
		14.5	11.4	9.4	7.4	5.2	3.0	1.3	0.5
20	19.1	14.8	12.5	9.8	7.4	5.0	3.1	1.0	0.4
		14.5	12.2	9.9	7.3	5.2	3.2	0.9	0.4

Una vez ingresados los datos el usuario tiene la posibilidad de ver diversos gráficos de dispersión de variables originales como de variables transformadas (ver figura 1A), con el fin de tener una idea de la relación existente entre ellas. Concluido el paso anterior el usuario ya puede construir e ingresar su

modelo (de volumen o ahusamiento) para su posterior ajuste por medio del método de mínimos cuadrados ordinarios.

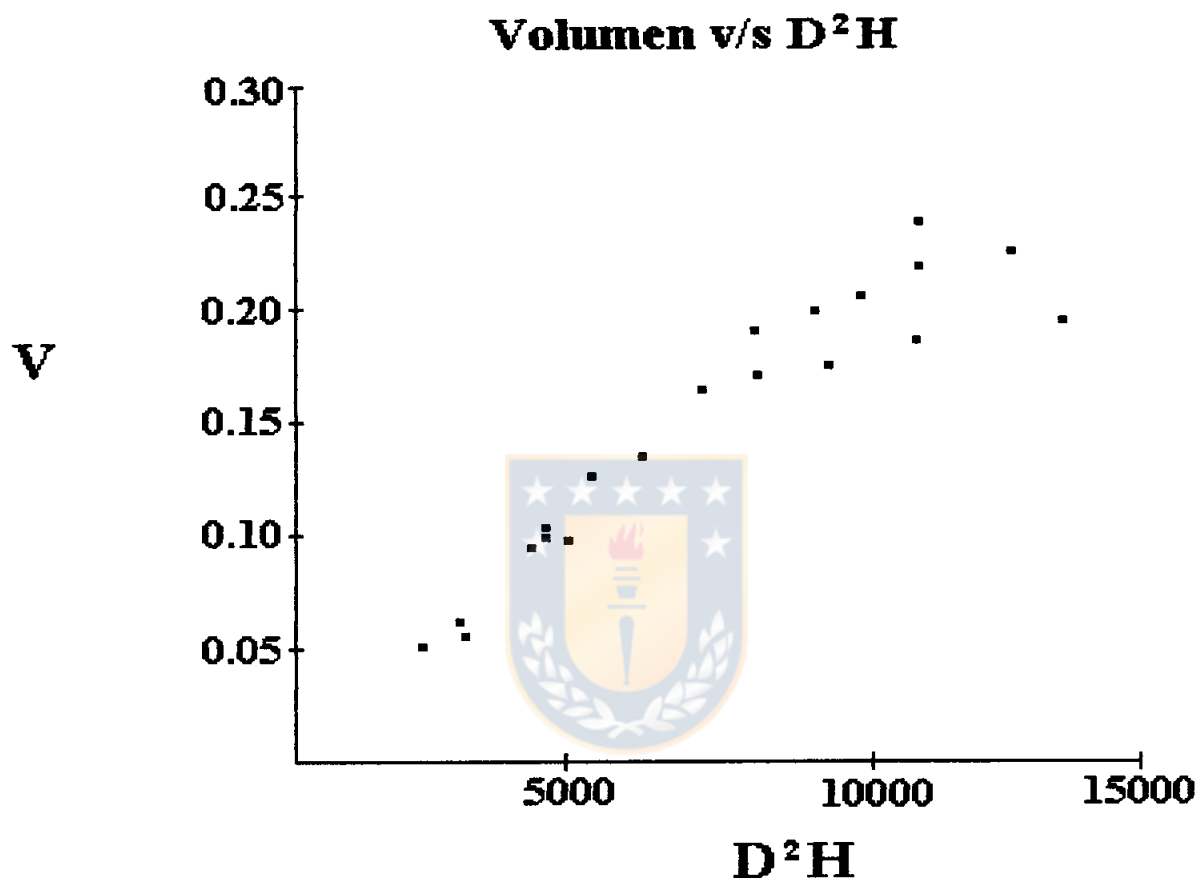


FIGURA N° 1A. Gráfico de dispersión de datos.

Para el ejemplo en cuestión se ingresó la siguiente función de volumen:

$$V = a + b H + c D + d D^2H$$

donde:

V = volumen ( $m^3$ ).

H = altura total (m).

D =  $D_{ap}$  (cm).

a, b, c, d = coeficientes de regresión.

Luego del ajuste del modelo, el usuario tiene la posibilidad de visualizar los errores desprendidos del ajuste de dos formas distintas, estas son:

i) En forma tabular; donde aparece el valor real, el valor estimado y el error para cada observación y,

ii) En forma gráfica; donde se grafica los errores versus los valores estimados de la variable dependiente, en este caso el volumen (ver figura 2A).

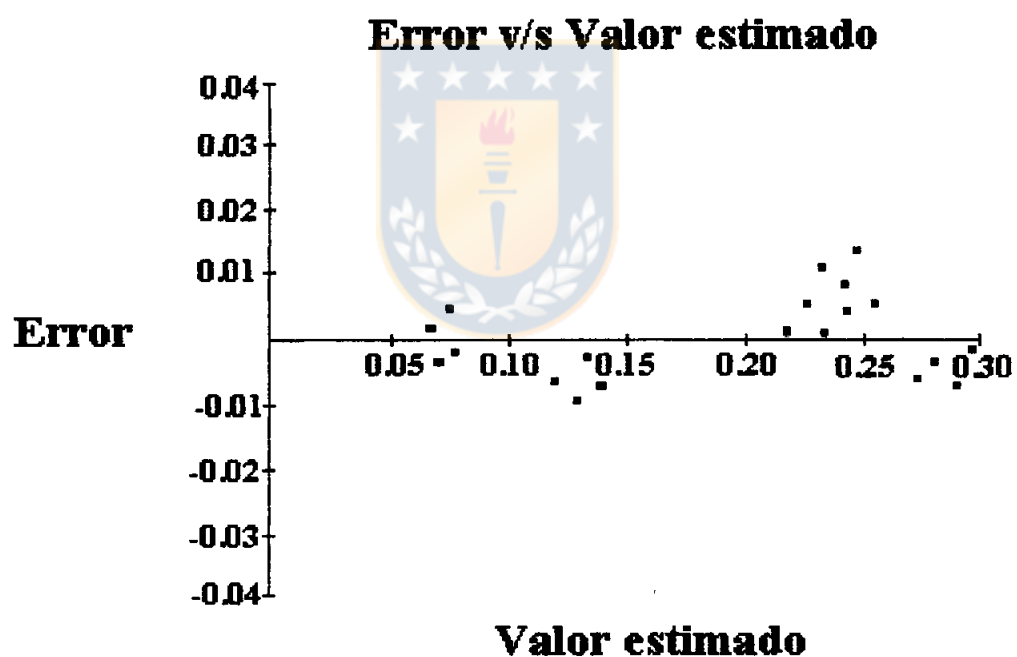


FIGURA N° 2A. Gráfico de dispersión de errores versus valor estimado.

Teniendo ingresados los datos, el modelo y los resultados del ajuste, se puede ingresar a la rutina de validación de supuestos. Esta rutina presenta en un menú todos los supuestos de regresión, así entonces, el usuario podrá validar los

supuestos en forma iterativa (uno a la vez, ver apéndices 2 al 6), para finalizar con la creación, almacenamiento e impresión del cuadro resumen, el que contiene la información más relevante de todo el proceso realizado. Para el ejemplo en cuestión se tiene la siguiente salida de resultados:

TABLA 2A: CUADRO RESULTADO PARA EL EJEMPLO.

### RESUMEN

<b>Nombre Predio</b>	: Los Peumos	<b>Superficie</b>	: 150 ha
<b>ROL S.I.I.</b>	: 1425-07	<b>Especie</b>	: Pino radiata
<b>Comuna</b>	: Chanco	<b>Edad</b>	: 18 años
<b>Provincia</b>	: Cauquenes	<b>Sitio</b>	: --
<b>Región</b>	: VII	<b>Origen</b>	: Plantación
<b>Nº Árboles</b>	: 20	<b>Archivo Datos</b>	: C:\VB\PEUMOS.DAT

$$V = - .356564 + .01272833H + .0216296D - .00001216D2H$$

#### Tabla ANDEVA

Fuente	Gr.Lib.	Suma Cuadrados	Med.Sum.Cuadrados	F-muestral	R2
Regresión	3	.107811	.035937	128.369	.960111
Error	16	.004479	.00028		.952631
Total	19	.11229			.016732

<b>Normalidad</b>	: Cumple	<b>Multicolinealidad</b>	: Cumple
<b>Homocedasticidad</b>	: No Cumple	<b>Especificación</b>	: No Cumple
<b>Autocorrelación</b>	: No Cumple		



APENDICE N°2: Interfaz resultado para el supuesto de normalidad.

**Enunciado :** Los errores poseen una distribución normal, cuyo promedio y varianza están dados por los supuestos 1 y 2.

$H_0: \hat{b}_1 = 0$  y  $\hat{b}_2 = 3 \Rightarrow$  Los errores están normalmente distribuidos.

$H_a$ : Los errores no están normalmente distribuidos.

$\alpha = 10\%$         $\alpha = 5\%$         $\alpha = 1\%$

$$\text{El estadístico es: } K^2 = n \left[ \frac{\hat{b}_1^2}{6} + \frac{(\hat{b}_2 - 3)^2}{24} \right]$$

$$\text{donde: } \hat{b}_1 = \frac{e_3^2}{e_2^2} \text{ y } \hat{b}_2 = \frac{e_4}{e_2^2} \text{ con: } e_r = \frac{1}{n} \sum_{i=1}^n e_i^r$$

$r = 2, 3, 4$  (momentos de orden 2, 3, 4.)

**RESULTADO**

El valor de  $K^2$  es 2.63

El valor crítico es 4.61

Como  $K^2$  es menor que el valor crítico, acepte  $H_0$ , luego no existe evidencia significativa que indique que la distribución de los errores no sea normal.

APENDICE N°3: Interfaz resultado para el supuesto de homocedasticidad.

**Enunciado :**

La varianza condicional de los errores, es constante u homocedástica.

$$H_0: \sigma_i^2 = \sigma_{i-1}^2 \quad i=2,3,\dots,n$$

$$H_a: \sigma_i^2 > \sigma_{i-1}^2$$

$\alpha = 10\%$

$\alpha = 5\%$

$\alpha = 1\%$

El estadístico es :

$$P = \left[ \frac{6n}{n^2 - 1} \right]^{1/2} \left[ \frac{\sum_{i=1}^n i e_i^2}{\sum_{i=1}^n e_i^2} - \frac{n+1}{2} \right]$$

**RESULTADO**

El valor de P es 0.957

El valor crítico es 1.282

Como P es menor que el valor crítico, acepto  $H_0$ , luego no existe evidencia significativa que indique que la varianza condicional de los errores no sea homocedástica.

APENDICE N°4: Interfaz resultado para el supuesto de autocorrelación.

**Enunciado :** No existe autocorrelación entre los errores.

**H<sub>0</sub> :** No existe autocorrelación entre los errores.

**H<sub>a</sub> :** Existe autocorrelación entre los errores.

alfa = 10%       alfa = 5%       alfa = 1%

**Autocorrelación de orden :** \_\_\_\_\_

Orden 1       Orden 2       Orden 3       Orden 4

El estadístico es :

$$Q = n \sum_{k=1}^p \hat{r}_k^2 \quad \text{con :} \quad \hat{r}_k^2 = \frac{\sum_{i=k+1}^n e_i + e_{i-k}}{\sum_{i=1}^n e_i^2}$$

donde :  $p$  = orden de autocorrelación presumido (v.g. 2, 3, 4).

**RESULTADO**

El valor de Q es 1.58

El valor crítico es 2.71

Como Q es menor que el valor crítico, acepto H<sub>0</sub>, luego no existe evidencia significativa que indique la existencia de autocorrelación entre los errores.

APENDICE N°5: Interfaz resultado para el supuesto de multicolinealidad.

**Enunciado :**

**No existe multicolinealidad entre las variables explicativas.**

**H<sub>0</sub> :** La variable independiente X<sub>i</sub> es no colineal con las restantes variables.

**H<sub>a</sub> :** La variable independiente X<sub>i</sub> es colineal con las restantes variables.

Alfa 10%

Alfa 5%

Alfa 1%

Con ayuda de regresiones auxiliares y dada la relación existente entre F y R<sup>2</sup>, se tiene el siguiente Estadístico :

$$F_i = \frac{R_{x_i \cdot x_2 \cdot x_3 \dots x_k}^2 / (k - 2)}{(1 - R_{x_i \cdot x_2 \cdot x_3 \dots x_k}^2) / (n - k + 1)}$$

donde : n = tamaño de la muestra.

k = número de var. indep. incluyendo el intercepto.

$R_{x_i \cdot x_2 \cdot x_3 \dots x_k}^2$  = coef. determinación de X<sub>i</sub> en las restantes variables X's.

**RESULTADO**

El valor F<sub>i</sub> es 15.96

EL valor crítico es 59.86

Como F<sub>i</sub> es menor que el valor crítico, acepto H<sub>0</sub>, luego no existe evidencia significativa que indique que la variable X<sub>i</sub> sea colineal con las restantes.

APENDICE N°6: Interfaz resultado para el supuesto de especificación.

**Enunciado :**

**El modelo de regresión está correctamente especificado,  
no existe error de especificación.**

**H<sub>0</sub> : El modelo está correctamente especificado.**

**H<sub>a</sub> : El modelo no está correctamente especificado.**

Alfa = 10%

Alfa = 5%

Alfa = 1%

**El estadístico es :**

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

**Durbin - Watson.**

**RESULTADO**

**El valor d es 1.98**

**Los valores críticos son  $1.5 < d < 2.5$**

**Como d está entre los valores críticos, acepto H<sub>0</sub>, luego el modelo está correctamente especificado.**



IX ANEXOS.

ANEXO N°1: Modelos de volumen más usados.

<u>Modelo.</u>	<u>Var. depend.</u>	<u>Var. indepen.</u>
1. $V=a+D^2$	V	$D^2$
2. $V=a+bD^3$	V	$D^3$
3. $\ln V=a+b\ln D$	$\ln V$	$\ln D$
4. $\ln V=a+b(1/\ln D)$	$\ln V$	$1/\ln D$
5. $V=a+bD+cD^2$	V	$D, D^2$
6. $V=a+bD^3+cD^{1/2}$	V	$D^3, D^{1/2}$
7. $V=a+bD+cD^{1/2}$	V	$D, D^{1/2}$
8. $V=a+b\ln D+c/D^2$	V	$\ln D, D^2$
9. $V=a+b\ln D+c/D^{1/2}$	V	$\ln D, 1/D^{1/2}$
10. $V=a+bD^2H$	V	$D^2H$
11. $V=a+bD+cD^2+dD^3$	V	$D, D^2, D^3$
12. $V=a+bD^3+c(D^2/D+10)$	V	$D^3, D^2, D$
13. $\ln V=a+b\ln D+c\ln H+d\ln F$	$\ln V$	$\ln D, \ln H, \ln F$
14. $V=a+bD^2HF$	V	$D^2HF$
15. $V=a+bH+cD^2+dD^2H$	V	$H, D^2, D^2H$

donde:

V= volumen ( $m^3$ ).

D= diámetro altura pecho (cm).

H= altura total (m).

F= clase de forma de Girard.

a, b, c, d = coeficientes de regresión.

ANEXO N°2: Modelos de ahusamiento más usados.

<u>Modelo.</u>	<u>Var. depend.</u>	<u>Var. indepen.</u>
1. $Y^2=a+bX+cX^2$	$Y= d/D$	$X= h/H$
2. $Y^2=a(X-1)+b(X^2-1)$	$Y= d/D$	$X= h/H$
3. $Y=a+bX+cX^2+dX^3+eX^4+fX^5$	$Y=d/D$	$X= h/H$
4. $Y=aX+bX^2+cX^3$	$Y= d/D$	$X=(H-h)/(H-1.3)$
5. $Y^2=aX+bX^2+cX^{10}+dX^{13}+eX^{20}$	$Y= d/D$	$X=(H-h)/(H-1.3)$
6. $Y=a(X^3-X^2)+b(X^8-X^2)+c(X^{40}-X^2)$	$Y= (d/D)^2-X^2$	$X=(H-h)/(H-1.3)$
7. $Y^2=aX^{3/2}+b(X^{3/2}-X^3)D+c(X^{3/2}-X^3)H$ $d(X^{3/2}-X^{32})DH+ e(X^{3/2}-X^{32})H^{1/2}$ $f(X^{3/2}-X^{40})H^2$	$Y=d/D$	$X=(H-h)/(H-1.3)$
8. $Y=aX^{R/2}$	$Y= d$	$X= (H-h)/H$
		$R= 1+ (H-h)/(H-1.3)$

donde:

d = diámetro fustal sin corteza a una altura h del árbol (cm).

D = diámetro altura de pecho (cm).

h = altura de medición de los diámetros fustales (m).

H = altura total (m).

a, b, c, d, e, f = coeficientes de regresión.



ANEXO N°3: Los supuestos de regresión y la prueba estadística para validarlos.

Supuesto N° 1: Normalidad de los errores.

Enunciado: Los errores poseen una distribución normal, cuyo promedio es cero y varianza es constante.

Simbólicamente:  $e \approx N(0, \sigma^2)$

Estadístico: Para validar este supuesto se utilizó la prueba  $K^2$  de Bowman y Shenton (Marshall et al., 1995), ésta se fundamenta en la propiedad de simetría que posee la distribución normal y en su eficiencia con muestras pequeñas.

El estadístico es:

$$K^2 = n \left( \frac{b_1}{6} + \frac{(b_2 - 3)^2}{24} \right) \quad \text{con} \quad b_1 = \frac{e_3^2}{e_2^3} \quad \text{y} \quad b_2 = \frac{e_4}{e_2^2}$$

donde: 
$$e_r = \frac{1}{n} \sum_{i=1}^n e_i^r$$

con  $r = 2, 3, 4$  (momentos de orden 2, 3, 4).

Hipótesis:  $H_0: b_1=0 \quad \text{y} \quad b_2=3 \Rightarrow$  Los errores están normalmente distribuidos.

$H_a:$  Los errores no están normalmente distribuidos.

Supuesto N° 2: Homocedasticidad de varianza.

Enunciado: La varianza condicional de los errores es constante u homocedástica.

Simbólicamente:

$$E(e_i^2) = \sigma^2 \quad i = 1, 2, \dots, n$$

Estadístico: Para validar este supuesto se utilizó la prueba  $Q'$  de Szroeter (Marshall et al., 1995).

El estadístico es:

$$Q' = \left( \frac{6n}{n^2 - 1} \right)^{1/2} * \left( \frac{\sum_{i=1}^n ie_i^2}{\sum_{i=1}^n e_i^2} - \frac{n+1}{2} \right)$$

Hipótesis: Ho:  $\sigma_i^2 = \sigma_{i-1}^2$

Ha:  $\sigma_i^2 > \sigma_{i-1}^2 \quad i = 1, 2, \dots, n$

Supuesto N° 3: Autocorrelación de errores.

Enunciado: No existe autocorrelación entre los errores.

En símbolos:

$$E(e_i, e_j) = 0 \quad i \neq j$$

Estadístico: Para validar este supuesto se utilizó la prueba  $Q$  de autocorrelación.

El estadístico es:  $Q = n \sum_{k=1}^o r_k^2$

con:  $r_k^2 = \frac{\sum_{i=k+1}^n e_i e_{i-k}}{\sum_i e_i^2}$

donde: o = orden de autocorrelación presumido.

Hipótesis: Ho: No existe autocorrelación entre los errores.

Ha: Existe autocorrelación entre los errores.

Supuesto N° 4: Multicolinealidad de variables independientes.

Enunciado: No existe multicolinealidad entre las variables independientes.

Estadístico: Utilizando regresiones auxiliares para luego correr una prueba "F" y contrastar las hipótesis.

El estadístico es:

$$F = \frac{R^2_{x1..x2x3...xk} / (k - 2)}{(1 - R^2_{x1..x2x3...xk}) / (n - k + 1)}$$

donde:

n= tamaño de la muestra.

k= n° de var. independientes incluido el inter-secto.

$R^2_{x1..x2x3...xk}$  = coeficiente de determinación de  $X_t$  en las restantes variables  $X'$ s.

Hipótesis:  $H_0$ : La variable independiente  $X_t$  es no colineal con las restantes.

$H_a$ : La variable independiente  $X_t$  es colineal con las restantes.

Supuesto N° 5: Especificación del modelo.

Enunciado: El modelo de regresión está correctamente especificado.

Estadístico: Para validar este supuesto se utilizó la prueba "d" de Durbin-Watson (Marshall et al., 1995).

El estadístico es:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

Hipótesis:  $H_0$ : El modelo de regresión está correctamente especificado.  
 $H_a$ : El modelo de regresión no está correctamente especificado.

