Autores:
Patricio Corvalán Vera
Jaime Hernández Palma

1.2 Requerimientos estadísticos en Dendrometría

1.2.1. Conceptos iniciales

La mensura forestal involucra el uso de las matemáticas, por lo cual el estudiante debe tratar los números de tal forma que le sirvan para describir los árboles y el bosque, o para que describan eficientemente todo tipo de cambios y relaciones en ellos, como por ejemplo el cambio en volumen con el paso del tiempo.

Por supuesto, el estudiante debe estar siempre alerta acerca de la dimensión y significado de los números que está usando. Por ejemplo, decir que una plantación de pino insigne en edad de cosecha en buenos sitios tiene un volumen total de 500 m^3/ha es una cifra razonable, pero decir que tiene 5.000 m^3/ha es, además de incorrecto, prácticamente imposible. Otro ejemplo es que un crecimiento de 4,0 m^3/ha/año es una productividad típica de bosque nativo y no de plantaciones de pino o eucalipto.

Uno de los objetivos que se busca lograr sobre los estudiantes es que, por una parte, sean capaces de descubrir errores de magnitud en las cifras descriptivas y que, por otra, manejen cifras típicas para los diferentes tipos de bosques. Además de la magnitud del número, es importante saber que tipo de número estamos analizando. A continuación se revisan los términos de mayor uso en mensura forestal.

Parámetro (*parameter*)

Un parámetro es una característica de una población que frecuentemente se puede expresar en forma numérica. Algunos ejemplos de parámetros son:

- Altura de los árboles
- Volumen total
- Área de bosque

Medición (*measurement*)

Una medición es una lectura o un valor asignado a un parámetro, generalmente obtenido con un instrumento. Ejemplos de mediciones son:

- Largo de una troza, medido con cinta de distancia
- Altura de un árbol, medido con un hipsómetro
- El Dap de un árbol medido con forcípula.

Estimación (*estimate*)

Una estimación es la asignación de un valor a un parámetro usando métodos indirectos, a menudo con la ayuda de mediciones. Una estimación se puede basar, por una parte, en el hecho de que aquella parte de que se midió (muestra) es representativa del total (población), y por otra, en que en modelo descriptivo (ejemplo: ecuación) de una combinación de mediciones puede establecer la relación cuantitativa entre dichas mediciones y el parámetro, y por lo tanto estimarlo. Ejemplos de estimaciones son:

- El *Dap* promedio de los árboles de un rodal podría ser estimado midiendo el 5 % de los árboles.
- El volumen del fuste de un árbol podría ser estimado asumiendo que la tiene forma de cono. Así, se usaría el *Dap* y la *H* para definir la base y altura del cono. De esta forma el volumen estimado del fuste es: $v = 1/3\ (\pi/4)\ Dap^2\ H$ (volumen de un cono).

Error (*error*)

En general, los errores pueden ser de tres tipos:

a) Humanos y erráticos: son errores que pueden ser evitados, por ejemplo el equivocarse en leer el instrumento, o usarlo en forma incorrecta.
b) Sistemáticos: son errores que afectan las mediciones en forma regular y son generalmente predecibles. Por ejemplo: usar una cinta de distancia que para la distancia marcada como 20 m, la distancia real es de 18,5 m. Todas las mediciones hechas con esta cinta tendrán siempre el mismo error.
c) Aleatorios: son errores irregulares en su magnitud y signo (+/-). Se originan por muchas causas, difíciles de determinar, y en general tienden a tener esperanza igual a cero. Es decir, la suma de todos los errores es igual a cero.

Exactitud (*accuracy*)

Es una medida de la diferencia entre el valor "verdadero" y el valor "medido" o "estimado". El valor verdadero puede ser determinado sólo efectuando medidas muy cuidadosas y usando instrumentos de alta precisión. La exactitud de una medición sólo puede ser conocida si el valor medido es conocido. En ciencias forestales, la exactitud de los instrumentos se revisa por comparación de estándares conocidos. Por ejemplo, se puede determinar la exactitud de un hipsómetro efectuando medidas estructuras con altura conocida (edificios, postes de luz, etc.)

Precisión (*precision*)

Es el mínimo grado en que una medida o estimación puede ser efectuada y depende de la calibración del instrumento, de la variación que existe en la población y del hecho que una estimación se basa en un supuesto predefinido. Ejemplos:

- La precisión de una medida de longitud efectuada con regla es, en general, de 0,5 mm.
- La precisión del *Dap* de un árbol efectuada con forcípula es, en general, de 0,5 cm.
- La precisión de la altura de un árbol efectuada con hipsómetro es, en general, de 0,5 m.

Como regla, se establece que la precisión de un instrumento es igual a la mitad de su mínima graduación, que es hasta donde el lector puede hacer una lectura confiable.

Sesgo (*bias*)

El sesgo en una medición o estimación se refiere a la presencia de un error sistemático que afecta todas las medidas o estimaciones en la misma manera. Ejemplos:

- Largos medidos con una cinta de distancia de 20 m que se rompió y fue unida nuevamente, con lo cual se perdieron 50 cm en total. Por lo tanto, cada vez que se miden 20 metros se comete un error (sistemático) de –50 cm.

- Si todas las lecturas son llevadas al entero completo (ejemplo: 5,4→5,0 y 5,9→5,0), entonces el total y el promedio de las medidas estarán sesgados.

El sesgo es medido mediante la diferencia entre el valor “real”, o estimación del valor “real”, y el valor obtenido con la medición. Si el sesgo es consistente, puede ser corregido y su efecto es removido. Si el sesgo es pequeño, en comparación a la precisión, entonces puede considerarse despreciable y no es necesario efectuar ningún tipo de ajuste.

Valor esperado (*expected value*)

El concepto de valor esperado tiene un significado especial cuando se usa en un sentido estadístico. En general podemos definir valor esperado como el resultado que se espera obtener si se repite la prueba muchas veces. De esta forma, el valor esperado de una muestra es el promedio poblacional (μ).

$$\mu = E(x) = \sum_{i=1}^{N} (x_i * p_i)$$

y el total de la población T es:

$$T = E(\frac{x_i}{p_i}) = \frac{1}{N} \sum_{i=1}^{N} (\frac{x_i}{p_i})$$

donde,

donde,	μ	= promedio de la población.
	$E(x)$, $E(x_i/p_i)$	= valores esperados
	x_i	= valor de la i^{th} unidad en la población
	N	= número de unidades en la población
	p_i	= probabilidad de ocurrencia del valor de x_i

Población (*population*)

Desde el punto de vista estadístico, una población es un conjunto de unidades individuales que usualmente se forma, usando criterios definidos, para efectuar algún tipo de medición. Ejemplo de poblaciones son:

- Todos los árboles de un sector del bosques o rodal.
- Todos los “robles” de un rodal.
- La cantidad de trozas en las canchas de acopio de un aserradero.
- Todas las semillas de un saco.
- Todas las ciudades de más de 50.000 habitantes de Chile.

Las poblaciones pueden pequeñas o grandes, restringidas a un área pequeña o a una gran superficie, homogéneas o heterogéneas.

Muestra (*sample*)

Desde el punto de vista estadístico, una muestra es una <u>parte</u> representativa de toda una población. Una muestra pequeña, de unos pocos individuos, puede ser adecuada para representar a una población homogénea, donde una muestra grande será necesaria para representa una población heterogénea.

1.2.2 Conceptos complementarios

A menudo se hará referencia a dos conceptos complementarios, de amplio uso en ciencias forestales: sistemas y modelos. A continuación se revisan ambos conceptos y se dan algunos ejemplos forestales.

1.2.2.1 Sistemas *(systems)*

Un sistema real es una fuente de datos del comportamiento de alguna parte del mundo real por la que mostramos interés. Esa parte de la realidad estará formada por un conjunto de elementos o componentes que interaccionan para alcanzar un objetivo común.

Los elementos poseen ciertas características o atributos (parámetros y variables) que toman valores numéricos o lógicos, y en conjunto, se denominan variables descriptivas del sistema. Por ejemplo, un sistema puede ser un rodal, en donde las variables descriptivas serán, las especies presentes, la cantidad de individuos, sus tamaños, etc.

Los sistemas tienen relaciones internas y externas. Las primeras conectan los elementos dentro del sistema; las segundas conectan los elementos con el mundo exterior. En el ejemplo, la competencia entre los árboles del rodal sería una relación interna, mientras que la relación de la productividad, en términos de biomasa, con la cantidad de energía solar que llega por unidad de tiempo y de superficie, sería un ejemplo de relación externa.

Se pueden clasificar los sistemas de muchas formas, algunas de las cuales son:

- Cerrados o abiertos. Se relaciona o no con el medio que lo rodea.
- Naturales o artificiales. Provenientes de la naturaleza o de carácter antrópico.
- Dinámicos o estáticos. Cambian o no los atributos en función de alguna variable definida (ejemplo: cambio temporal).
- Estables o inestables. Regresan o no al equilibrio después de una perturbación.
- Estocásticos o determinísticos. Incluyen elementos aleatorios o no.
- Adaptativos o no adaptativos. Responden o no a cambios en el medio.
- Lineales o no lineales. Las relaciones entre elementos son de carácter lineal o no.

El primer paso en el estudio de un sistema es la construcción de un modelo que lo represente. Un modelo, desde el punto de vista científico, puede ser entendido como una abstracción de un sistema real. Dada la trascendencia del concepto se estudia su significado y diferentes enfoques en el siguiente punto.

1.2.2.2 Modelos *(models)*

Un modelo científico puede definirse como una representación simplificada de un sistema real o un proceso o una teoría, con el que se pretende aumentar la comprensión, hacer predicciones y, en algunos casos, ayudar a controlar el sistema.

Personas diferentes pueden entender en forma diferente el concepto de modelo. Una clasificación de los diferentes tipos de modelos se presenta a continuación:

- Modelos físicos: Son representaciones de sistemas físicos y están descritos por variables medibles. Algunos ejemplos son: aviones a escala, mapa de relieve (imitación), reloj de arena (analógico) o un nuevo circuito eléctrico (prototipo).

- Modelos mentales: Son modelos heurísticos o intuitivos que sólo existen en nuestras mentes. El ser humano tiene la habilidad de acumular experiencia que usa par construir sus modelos de comportamiento y razonamiento.

- Modelos simbólicos: Son aquellos que incluyen operaciones matemáticas o lógicas y pueden ser utilizados para formular una solución a un problema. Son más económicos que los modelos físicos y de mayor facilidad de construcción. Se subdividen en modelos matemáticos y no matemáticos. En mensura forestal se recurre habitualmente a modelos matemáticos que predicen las variables descriptivas de interés para árboles individuales o para rodales completos. Ejemplos de modelos usados en ciencias forestales son:

Modelo de perfil fustal (dendrométrico): predice el diámetro *d c/c* a cualquier altura *h* en el fuste, en función del diámetro a la altura del pecho *D* y la altura total *H*.

$$d = D\ (a + b * \left(\frac{h}{H}\right) + c * \left(\frac{h}{H}\right)^2)$$

1

donde: a, b, c = parámetros de la ecuación

Modelo predictor de la altura dominante (dasométrico): predice la altura dominante[1] del rodal *H* en función de la edad *E* del mismo.

$$\ln H = a + b\left(\frac{1}{E}\right)$$

donde: a, b = parámetros de la ecuación

La construcción de este tipo de modelos requiere de observaciones hechas sobre el sistema real (mediciones). Es necesario observar las relaciones entre las variables descriptivas involucradas para lo cual es de utilidad el uso de gráficos que muestren los valores de una variable (por ejemplo: *x*) en función de otra (por ejemplo: *y*) y luego determinar el tipo de relación existente (por ejemplo: lineal, exponencial, logarítmica, etc.). La estimación de los parámetros de las ecuaciones puede efectuarse

[1] La altura dominante se define como la altura promedio de todos los árboles dominantes del rodal.

mediante las técnicas estadísticas apropiadas (ajuste lineal por mínimos cuadrados, ajustes no lineales, etc.).

1.2.3. Requerimientos estadísticos básicos

A continuación se detallan los conocimientos mínimos que el estudiante de dendrometría debería poseer. Si se tuviese dudas respecto de alguno de ellos es necesario que revise la bibliografía estadística básica, en especial la recomendada al final del capítulo.

1.2.3.1. Estadísticas descriptivas

Las estadísticas descriptivas pueden ser calculadas para datos provenientes de una muestra aleatoria x_i, donde $i=1, ..., n$ o para datos agrupados en clases en donde x_i corresponde al valor de la clase, y f_i a las frecuencias de cada una ($i=1, ..., k$). En este último caso, el número total de observaciones esta dado por:

$$n = \sum_{i=1}^{k} f_i$$

A continuación se revisan las principales estadísticas descriptivas para una variable aleatoria, y en los casos necesarios se entregan las fórmulas para datos agrupados y no agrupados.

Variable aleatoria

Considérese un espacio muestral *S* sobre el cual se tiene definida un función que establece con que probabilidad pueden ocurrir todos y cada uno de los valores contenidos. Si *X* es una función de valor real, definida para el espacio muestral, de manera que transforma el resultado de la muestra en puntos sobre la recta de los reales, entonces *X* es una variable aleatoria.

Función de probabilidad de una variable aleatoria

En el rango R_x una variable aleatoria continua presenta una determinada distribución *f(x)* de probabilidades de que cada valor ocurra. Esta función se denomina función de distribución de probabilidades *fdp* de *x* y satisface las siguientes condiciones:

$f(x) \geq 0$	Para todos los valores de *x* dentro del rango R_x.
$\int_{R_x} f(x)dx = 1$	
$f(x) = 0$	Si *x* no está en el rango R_x.

Media Aritmética o Promedio (*arithmetic mean / average*)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} f_i x_i$$ Para datos agrupados.

Media ponderada (*weighted mean*)

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$$

Cada valor x_i está asociado a un factor de peso $w_i \geq 0$.

peso total $= \sum_{i=1}^{n} w_i$

Media geométrica (*geometric mean*)

$$\bar{x}_{geom} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

$$\bar{x}_{geom} = \sqrt[n]{x_1^{f_1} \cdot x_2^{f_2} \cdot \ldots \cdot x_n^{f_n}}$$ Para datos agrupados.

Media armónica (*harmonic mean*)

$$\bar{x}_{harm} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

$$\bar{x}_{harm} = \frac{n}{\sum_{i=1}^{n} \frac{f_i}{x_i}}$$ Para datos agrupados.

Relación entre media aritmética, geométrica y armónica

$$\bar{x}_{harm} \leq \bar{x}_{geom} \leq \bar{x}$$

Moda (*mode*)

La moda de un conjunto de datos corresponde al valor que ocurre con la mayor frecuencia. La moda puede no existir y si existe puede no ser única.

Mediana (*median*)

Si los datos de la muestra de disponen en orden ascendente, la mediana estará dada por el valor en la posición *(n+1)/2*. Cuando el total de valores es impar, entonces la mediana está dada por el valor central de los datos ordenados. Cuando el número total es par, la mediana es calculada como el promedio entre los dos valores centrales.

Relación empírica entre Media, Mediana y Moda

Media – Moda = 3 (Media – Mediana)

Cuartiles (*quartiles*)

Si los datos son ordenados en forma ascendente, el $j^{ésimo}$ cuartil Q_j, para $j=1,2$ *y* 3, está dado por el valor en la posición $j(n+1)/4$. Podría ser necesario interpolar entre otros valores.

Deciles (*deciles*)

Si los datos son ordenados en forma ascendente, el $j^{ésimo}$ decil D_j, para $j=1,2,...,9$, está dado por el valor en la posición $j(n+1)/10$. Podría ser necesario interpolar entre otros valores.

Percentiles (*percentiles*)

Si los datos son ordenados en forma ascendente, el $j^{ésimo}$ percentil P_j, para $j=1,2,...,99$, está dado por el valor en la posición $j(n+1)/100$. Podría ser necesario interpolar entre otros valores.

Desviación estándar *(standard deviation)*

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Varianza *(variance)*

$$\mathrm{var} = s^2$$

Coeficiente de variación *(coefficient of variation)*

$$cv = 100 \cdot \frac{s}{\bar{x}} \qquad (\%)$$

Variable estandarizada (normalizada)

$$z = \frac{x_i - \bar{x}}{s}$$

Momentos de una variables aleatoria *(moments of a random variable)*

Son valores esperados de ciertas funciones de la variable aleatoria y constituyen una colección de medidas descriptivas que caracterizan a la *fdp*.

El *r-ésimo* momento alrededor del origen de una variable aleatoria está dado por:

$$m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$$

Para r = 0 se obtiene la media

El *r-ésimo* momento alrededor de la media de una variable aleatoria está dado por:

$$m_r = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^r$$

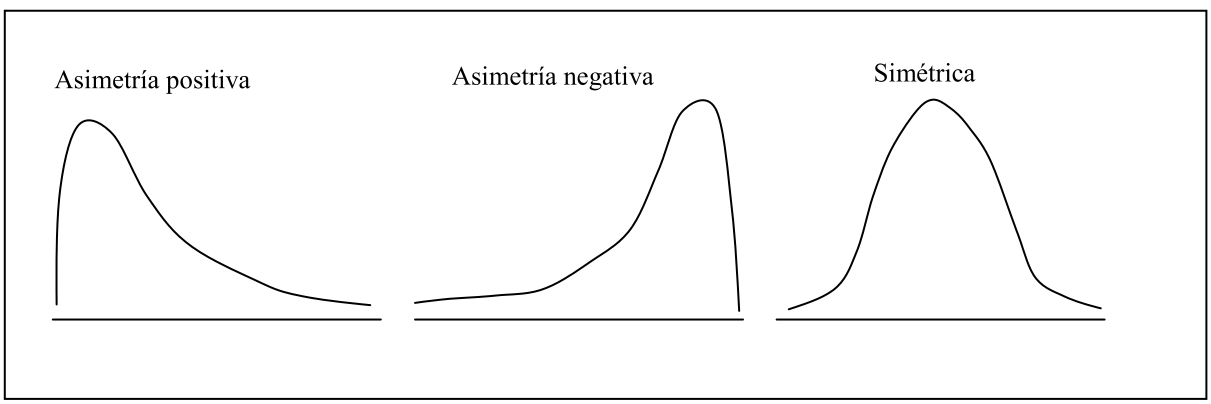
Para r = 0, m = 1
Para r = 1, m = 0
Para r = 2, m = varianza

Coeficiente de Sesgo *(coefficient of skewness)*

Este coeficiente define el grado de desvío de la función de densidad de probabilidad de una variable aleatoria respecto de la simetría.

$$\alpha_3 = \frac{m_3}{(m_2)^{(3/2)}}$$

Donde m_2 y m_3 son los momentos de segundo y tercer orden respecto de la media.

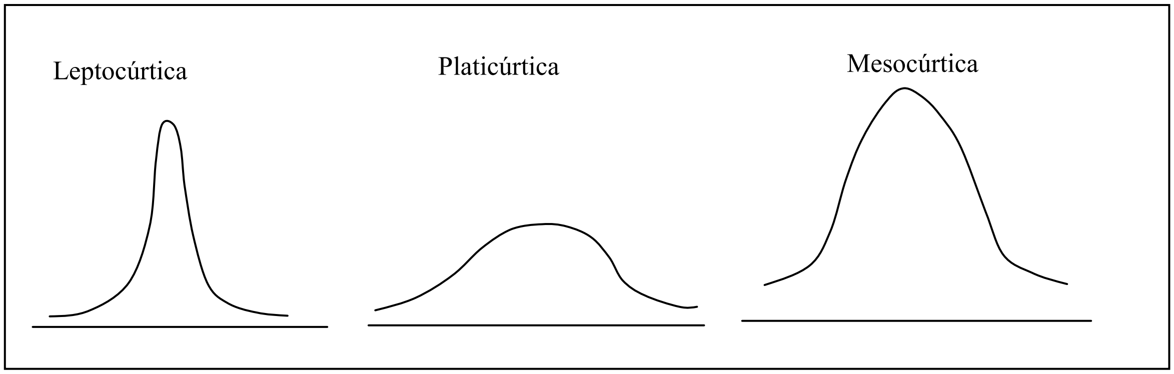


Coeficiente de Curtosis (coefficient of kurtosis)

La curtosis es el mayor o menor grado de achatamiento de la función de densidad de probabilidades. Mientras mayor sea el coeficiente de curtosis, menor es el achatamiento.

$$\alpha_4 = \frac{m_4}{(m_2)^2}$$

Donde m_2 y $m4$ son los momentos de segundo y cuarto orden respecto de la media.

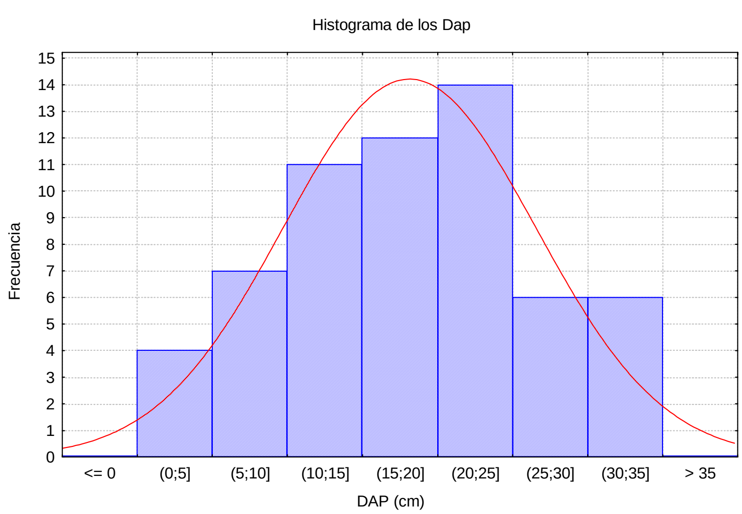


1.2.3.2 Exploración y representación gráfica de los datos

Un complemento a las estadísticas descriptivas, de gran utilidad en el análisis, es el uso de representaciones gráficas de las variables bajo estudio. Se puede observar el comportamiento de una sola variable, sus frecuencias por clases o su cercanía a una distribución normal. Por otra parte, se podría observar el grado de relación entre dos o más variables, definir el tipo de modelo de modelo matemático que mejor represente dicha relación, u observar la relación de tres variables en un gráfico tridimensional. Las utilidades de un análisis visual, exploratorio de los datos, son múltiples y dependen del grado de especialización del observador. A continuación se resumen algunas de dichas técnicas y se complementa la presentación con algunos ejemplos.

1.2.3.3 Histogramas y gráficos de frecuencia de datos por clase *(histograms)*

Los gráficos de frecuencias por clase son de utilidad en el estudio de una variable, en especial cuando se busca definir su similitud con una distribución de probabilidades teórica, como por ejemplo la normal.



Histograma para clases de Dap de 5 cm de ancho. Los datos usados están al final del subcapítulo.

1.2.3.4 Gráficos de dispersión de datos *(scatterplots)*

Los gráficos de dispersión de datos, o *ploteos,* son de gran utilidad cuando se busca observar la existencia de relación entre variables. Además permiten definir el tipo de relación y seleccionar modelos matemáticos que podrían ser útiles para representar dicha relación.

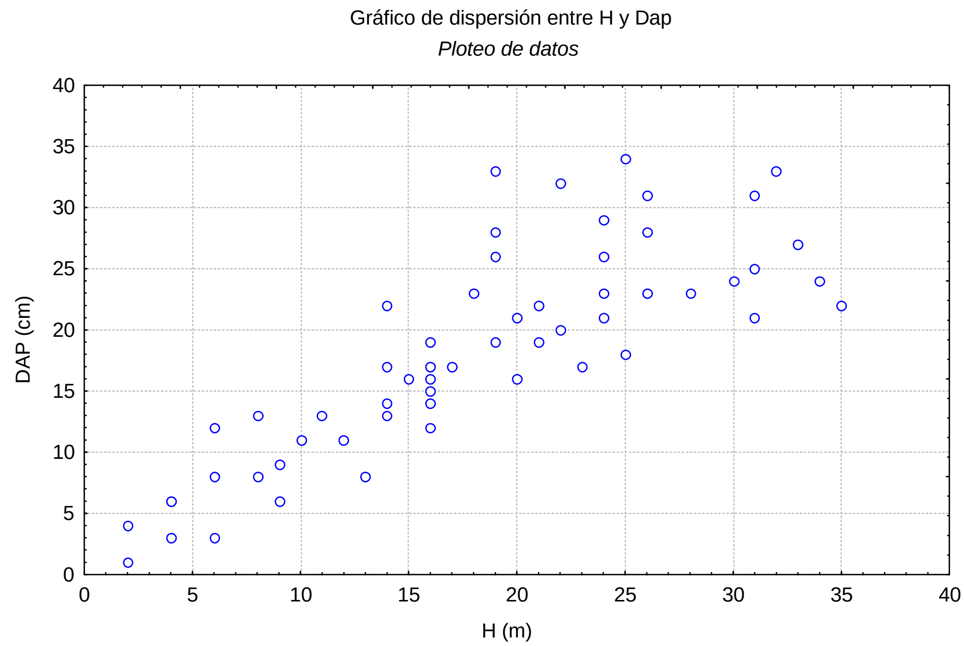
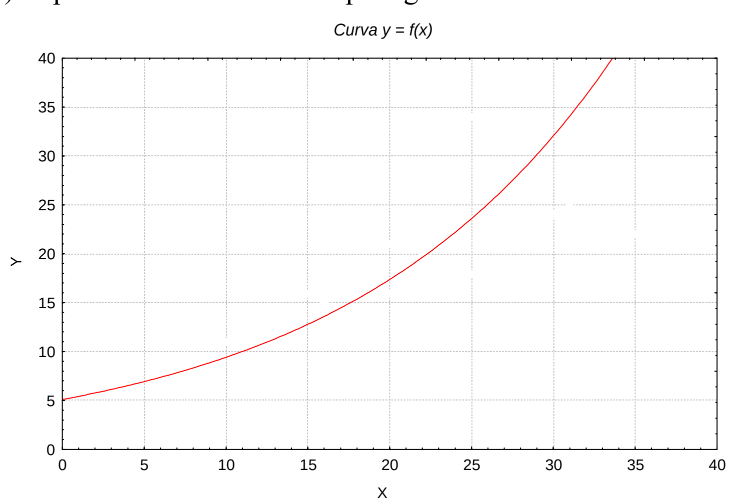


Gráfico de dispersión de valores de altura total (H) *versus* Dap. Los datos usados están al final del subcapítulo

1.2.3.5. Variables contínuas

Cuando la relación entre dos variables es conocida y además se tiene un modelo matemático que la representa ($y = f(x)$) se puede usar directamente para graficar la relación entre ambas variables.



Curva que representa la relación entre las variables *x* e *y*.

1.2.3.6 Modelación y ajuste de curvas

Los conceptos teóricos involucrados en la construcción de un modelo ya fueron expuestos en el punto 1.2. En esta sección se revisarán los aspectos prácticos en la construcción de un modelo matemático que describa la relación existente entre dos variables dendrométricas.

Una vez efectuado el análisis exploratorio de los datos, a través de estadísticas descriptivas y gráficos, es necesario seleccionar el o los modelos que serán usados.

Elección del modelo y linealización de curvas

Cuando la relación entre dos variables es de tipo lineal, es decir que un cambio de la variable independiente (predictora) significa un cambio proporcional fijo en la variable dependiente (predicha), el modelo queda definido mediante la ecuación:

$$y = a + b \cdot x$$

Este modelo, implica que para la población bajo estudio existe una relación lineal entre la variable x y la variable y. Esta relación, al igual que un parámetro poblacional, existe en la realidad pero para poder conocerla sólo podemos, a través de una muestra, estimar cuales serán los valores de los parámetros del modelo a y b. Para ello, existen diferentes aproximaciones metodológicas: métodos numéricos, algebraícos o estadísticos. Uno de los más populares es el ajuste por el método de los mínimos cuadrados, o regresión lineal por mínimos cuadrados.

Cuando la relación entre las variables no es lineal, en general, se pueden seguir dos caminos: linealizar el modelo que describe la relación o emplear métodos de ajuste no lineales.

Algunos ejemplo de curvas linealizables son:

$$y = a \cdot x^b \longrightarrow \ln y = \ln a + b(\ln x)$$

$$y = a \cdot e^{b \cdot x} \longrightarrow \ln y = \ln a + b \cdot x$$

$$y = \frac{1}{(a + b^{-c \cdot x})} \longrightarrow \textbf{\textit{No puede ser linealizada}}$$

Regresión lineal

Las fórmulas que siguen están definidas para un conjunto de n pares de datos del tipo $\{(x_i, y_i)\}$, para $i=1,..., n$. Los supuestos de una regresión normal son que las variables x son fijas y que las variables y son aleatorias que tienen una distribución normal de varianza igual a σ^2. Los supuestos de correlación normal son que los datos $\{(x_i, y_i)\}$ constituyen una muestra aleatoria extraída de una población normal bivariada.

Para una regresión de y sobre x: $E(y / x) = b_0 + b_1 \cdot x$

Donde *E(y/x)* es la media de la distribución de *y* para un *x* dado. Para una línea recta, ajustada por el método de los mínimos cuadrados, los valores de b_o y b_1 se obtienen mediante las ecuaciones normales, cuya solución es:

$$b_1 = \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2}$$

$$b_0 = \frac{\sum y_i}{n} - b_1 \frac{\sum x_i}{n} = \bar{y} - b_1 \cdot \bar{x}$$

Error estándar de la estimación:

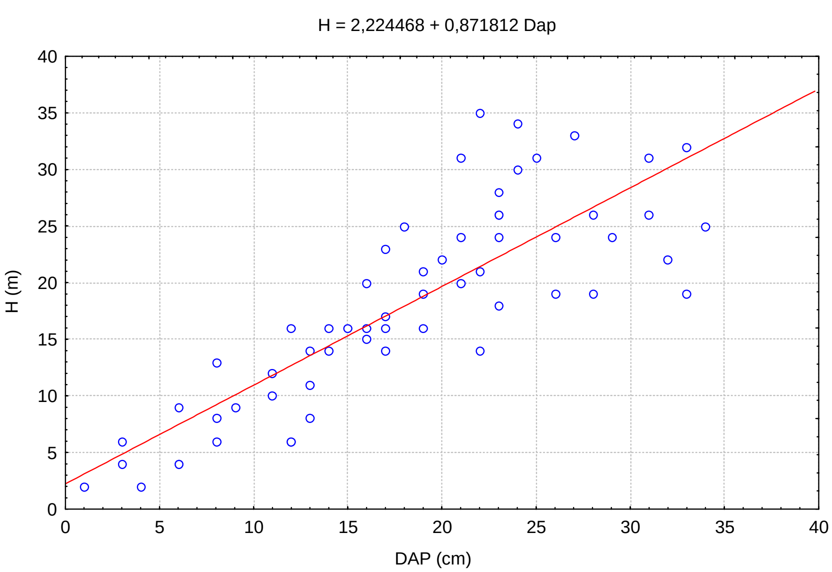
$$s_e = \sqrt{\frac{\sum \left[y_i - (b_0 + b_1 \cdot x) \right]^2}{n-2}}$$

1.2.3.7 Correlación

Una estimación del coeficiente de correlación ρ entre las variables está dada por:

$$r = \sqrt{\frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\left[\sum (x_i - \bar{x})^2\right]\left[\sum (y_i - \bar{y})^2\right]}}$$

A continuación se muestra el ajuste de la función H = a + b Dap, usando los datos al final del subcapítulo.



<u>Datos del ajuste</u>:

r = 0,83 significativa al 95% (α = 0,05)

r^2 = 0,67

error estándar del estimado = 4,97

estimador del intercepto= 0,224468
error estándar = 1,534

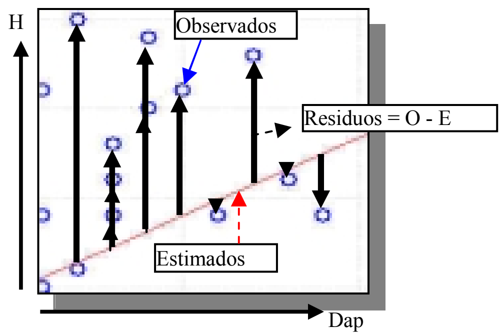
estimador de la pendiente = 0,871812
error estándar = 0,076

24468 + 0,871812 Dap)

1.2.3.8 Análisis de los residuos

En general no basta con realizar el ajuste, obtener los estimadores de los parámetros y los indicadores de la bondad del ajuste (correlación, error de la estimación, etc.). Es necesario efectuar un análisis de las desviaciones entre los valores observados (valores de la muestra) y los valores estimados (valores que entrega el modelo).

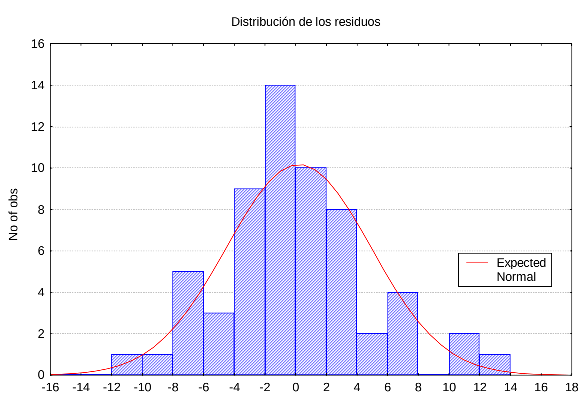
Por ejemplo, para un valor de *x* igual a 20, se <u>observó</u> que el valor de *y* es igual a 22, en la muestra de datos. Ahora, usando el modelo lineal ajustado para la relación entre ambas variables se obtiene que para un valor *x* igual a 14 el modelo <u>estima</u> un valor de *y* igual a 17,7. La diferencia entre el valor observado (20) y el valor estimado (17,7) se conoce con el nombre de error o <u>residuo</u>. En este caso, el residuo es negativo y tiene un valor de +2,3.



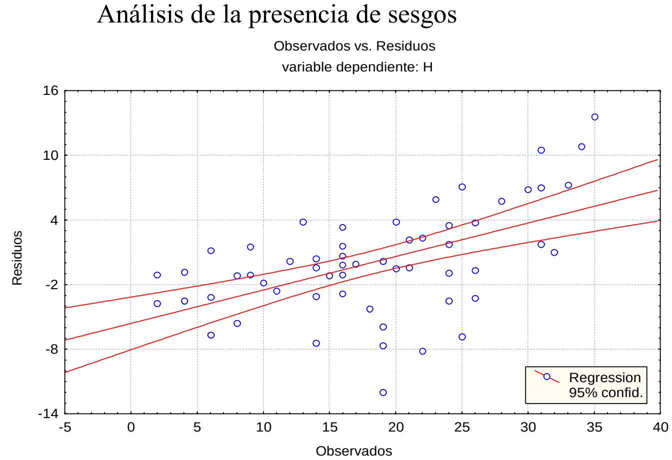
Esquema de la relación entre valores observados, predichos y los residuos.

Por último se presentan tres gráficos de utilidad en el análisis de los residuos. Por supuesto, el análisis debe ir más allá que sólo mostrar los gráficos, discutiendo y concluyendo en función de ellos.

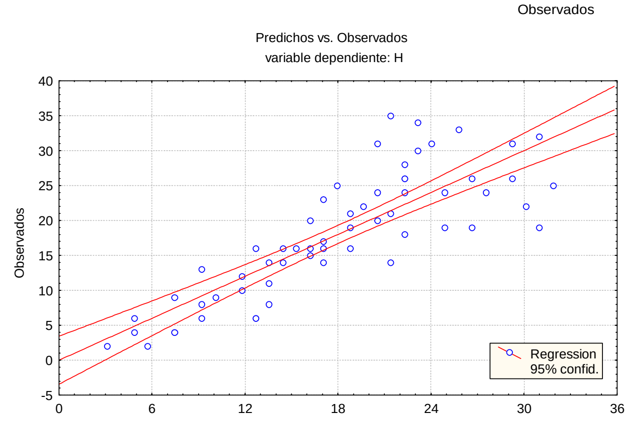
Normalidad de la distribución de los residuos



Análisis de la presencia de sesgos



Predichos *versus* observados



Datos usados en el ajemplo:

Rodal	Árbol	Dap (cm)	H (m)	Rodal	Árbol	Dap (cm)	H (m)
1	1	3	4	2	31	22	21
1	2	6	4	2	32	22	14
1	3	8	8	2	33	23	18
1	4	11	10	2	34	26	19
1	5	12	6	2	35	29	24
1	6	13	11	2	36	28	19
1	7	12	16	2	37	31	26
1	8	14	14	2	38	32	22
1	9	15	16	2	39	33	19
1	10	17	16	2	40	34	25
1	11	19	21	3	41	1	2
1	12	21	20	3	42	3	6
1	13	21	24	3	43	6	4
1	14	23	26	3	44	8	13
1	15	23	24	3	45	9	9
1	16	26	24	3	46	13	8
1	17	27	33	3	47	13	14
1	18	28	26	3	48	16	15
1	19	31	31	3	49	16	20
1	20	33	32	3	50	17	17
2	21	4	2	3	51	17	23
2	22	6	9	3	52	18	25
2	23	8	6	3	53	20	22
2	24	11	12	3	54	21	31
2	25	12	6	3	55	22	35
2	26	14	16	3	56	23	26
2	27	16	16	3	57	23	28
2	28	17	14	3	58	24	34
2	29	19	19	3	59	24	30
2	30	19	16	3	60	25	31

1.2.4 Exactitud de los cálculos

La mayoría de los números usados en mensura forestal son enteros o aproximaciones de números puros. Por ejemplo, si se miden 10 árboles en una muestra, el número 10 es un valor discreto o entero. En contraste, si estimamos el volumen de un árbol en 1,2 m^3, corresponde a una aproximación que implica que los límites dentro de los cuales se encuentra la estimación son: 1,15 y 1,25 m^3. La precisión de la aproximación se indica ya sea usando el denominador en una fracción o el número de dígitos significativos en el número decimal. Algunos ejemplos son:

- 23,5 cm implica que el valor se encuentra entre 23,45 y 23,55 cm
- 23½ cm implica que el valor se encuentra entre 23¼ y 23¾ cm

Un dígito significativo es cualquier dígito que denota el verdadero tamaño de la unidad; los dígitos significativos son aquellos que se leen de izquierda a derecha en un número, empezando desde el primer dígito distinto de cero hasta el último dígito, que puede ser cero. Una transformación de unidades en las cuales un número esta escrito no afecta el número de dígitos significativos. Ejemplo:

- 15,60 cm^2 = 0,001560 m^2

Ambos números tienen 4 dígitos significativos. Consecuentemente, el número de lugares decimales no es importante, pero el número de dígitos significativos indica la precisión de la aproximación. Para áreas basales menores a 1,0 m^2, los diámetros registrados redondeando al centímetro entero más cercano, sólo permiten un dígito significativo en la aproximación de la correspondiente área basal. Ejemplos para áreas basales:

- Los límites de un diámetro registrado como 10 cm son 9,5 y 10,5 cm, lo cual genera áreas basales entre 70,9 y 86,6 cm^2 o 0,00709 y 0,00866 m^2.

- Para un diámetro de 160 cm cuyos límites son 159,5 y 160,5 cm, con sus correspondientes áreas basales de 1,998 y 2,023 m^2, la aproximación correcta es 2,0 m^2.

En mensura forestal los cálculos deben indicar si los resultados son valores esperados para un gran número de estos valores involucrados o son estimaciones relacionadas a un número limitado de árboles.

En este último caso, el número de dígitos significativos debe ser realista. Es una práctica normal es redondear los datos a la unidad entera menor facilitando los cálculos posteriores. Se debe ser cuidadoso en asignar los números marginales al entero mayor y menor, alternativamente, para evitar sesgos.

El redondeo de cifras aumenta la varianza pero el monto del aumento no es importante. Ejemplo de redondeo de datos:

	Valor verdadero (cm)	**Valor redondeado (cm)**	
	7,35	7	
	8,82	9	
	9,50	10	Al entero mayor
	7,89	8	
	6,50	6	Al entero menor
	9,65	10	
	8,53	9	
	7,14	7	
	8,41	8	
	6,18	6	
Total	**79,97**	**80**	
Varianza	**1,42**	**2,22**	
Promedio	**8,00**	**8**	

A veces los registros son hechos a la unidad completa menor. Sino se corrige, el total y el promedio estarán sesgados.

El valor esperado del sesgo, para un gran número de casos, será igual a la mitad del valor de la unidad de medición. Ejemplo:

	Valor verdadero (cm)	Valor redondeado (cm)	Sesgo esperado
	7,35	7	-0,5
	8,82	8	-0,5
	9,50	9	-0,5
	7,89	7	-0,5
	6,50	6	-0,5
	9,65	9	-0,5
	8,53	8	-0,5
	7,14	7	-0,5
	8,41	8	-0,5
	6,18	6	-0,5
Total	**79,97**	**75**	**-5,0**
Varianza	**1,42**	**1,17**	**0**
Promedio	**8,00**	**7,5**	**-0,5**

1.2.5 Errores de medición y medición de la variación

Los errores de medición pueden ser causados por:

a) Errores humanos, como por ejemplo registrar 20 cm cuando en realidad son 30 cm, o aplicar incorrectamente una técnica de medición, o usar descuidadamente un instrumento.
b) Errores instrumentales, como por ejemplo usar una cinta de distancia que ha sido reparada y donde marca 20 m en realidad representa una distancia de 19 m.
c) Errores debido a la naturaleza del objeto que se mide, como por ejemplo los errores de la magnitud de la circunferencia cuando se miden árboles con fustes irregulares, con grandes contrafuertes.

Los errores de las dos primeras categorías deberían ser eliminados o al menos minimizados a través de un adecuado entrenamiento del personal en la mantención y procedimientos de medición. Los errores de la tercera categoría sólo pueden ser minimizados aplicando las técnicas más adecuadas para cada caso; éste es el trabajo del experto en dendrometría.

Los errores también pueden ser categorizados en:

a) Errores aleatorios: estos ocurren a ambos lados del valor verdadero (positivo y negativo) y su magnitud es aleatoria, en donde los errores pequeños son mucho más frecuentes que los errores grandes; esto es que tienden a tener un valor esperado igual a cero. Este tipo de errores no afecta el valor esperado de la media y por lo tanto tampoco afecta al total.
b) Errores sistemáticos: son aquellos que ocurren sólo a un lado del valor verdadero (positivo o negativo) y en una magnitud regular. Este tipo de error introduce sesgo en el valor esperado de la media y del total.

Ejemplos de errores:

- Errores aleatorios (simétricos): la diferencia entre mediciones sucesivas de la distancia entre dos puntos. El promedio de un número alto de mediciones de esta medida es insesgado.
- Errores aleatorios (sólo a un lado): los errores en los cálculos basados en que el área basal de un árbol es circular, lo cual generará un valor siempre mayor al área verdadera. Esto produce un sesgo irregular.

- Errores sistemáticos: medición de diámetros fustales con una forcípula con el brazo móvil defectuoso, que siempre aumenta la lectura en 0,5 cm. Lo anterior genera valores esperados de la media y del total sesgados (sesgo regular).

Análoga a la diferencia entre mediciones repetidas sobre el mismo objeto se tienen las diferencias entre mediciones del mismo atributo sobre diferentes individuos de una misma población. Generalmente un grupo reducido de individuos presentan valores muy pequeños, otro grupo reducido presentan valores muy grandes, y un gran número de individuos presenta valores cercanos al promedio de la población. Muchas poblaciones de seres vivos presentan lo que los estadísticos han denominado distribución normal alrededor del promedio o media. Una distribución normal resulta cuando las causas de la variación en tamaño son numerosas e independientes. Por ejemplo, los diámetros de los árboles de un bosque coetáneo[2] varían debido a que los árboles tiene distinto genotipo, crecen sobre diferentes condiciones locales de sitio (suelo, disponibilidad de agua, microclima, etc.). Como resultado, los diámetros varían y la distribución de las frecuencias de diámetros en la población puede ser predicha debido a que las diferencias aleatorias entre ellas, al igual que los errores aleatorios, se espera que tenga una distribución normal.

La variación se caracteriza a través de la medida estadística denominada desviación estándar, y de su cuadrado llamada varianza. En mensura forestal ambas medidas son generalmente estimadas a partir de muestras que reducen el costo de obtención de la información. Cada individuo es llamado unidad de muestreo y a cada uno de ellos, por ejemplo, se le asocia una medida de diámetro.

En una distribución normal se espera que alrededor del 67% de las observaciones se encuentre en el rango (μ - σ) a (μ + σ), y el 95 % en el rango (μ - 2σ) a (μ + 2σ), donde μ es la media poblacional y σ su desviación estándar, cuya expresión es:

$$s^2 = \sum_{i=1}^{n} \frac{(x_i - \overline{x})^2}{n-1}$$

Donde s^2 = varianza estimada a partir de *n* unidades muestrales.
i = unidades muestrales *1, 2, 3 ...., n*
x_i = valor de la i-ésima unidad muestral
$\overline{x}$ = media estimada a partir de *n* unidades muestrales

1.2.6 ¿Peso o volumen?

Las unidades que se usen para definir las diferentes partes del árbol y sus productos deben reflejar lo mejor posible el usos que se les dará a los mismos. Además de fines científicos, los árboles son medidos para:

- Obtener un valor de compra o venta.
- Obtener información útil para el manejo forestal acerca de la cantidad de madera acumulada y sus variación temporal (crecimiento o productividad).

[2] Bosque coetáneo es aquel en que todos los individuos pertenecen a la misma clase de edad. Por ejemplo, plantaciones artificiales de pino o renovales de *Nothofagus*.

Para ello, los parámetros de mayor uso son el peso y el volumen. Ambos son mutuamente deducibles si es que la densidad de la madera es conocida. A continuación se mencionan algunas de las ventajas y desventajas asociadas al uso de cada uno:

Mediciones en base a peso en verde (madera con agua)

Ventajas

- Rápido con bajo costo fijo por carga
- Confiable (automatizable)
- Medida directa (pesa)
- Se espera consistencia en medidas repetidas

Desventajas

- Las instrumentos de pesaje son caros
- Los instrumentos de pesaje no son utilizable en el bosque
- La presencia de agua en la madera es fuente de variación en las mediciones
- Es difícil predecir el crecimiento
- Sólo es operativo para medir el fuste

Mediciones en base Peso seco en horno secador (madera seca)

Ventajas

- Refleja la medida en que la fibra o pulpa se comercializa
- Pon énfasis en la valoración hecha por el productor

Desventajas

- Sólo puede ser estimada en forma indirecta, vía muestreo o predicción, a partir del peso en verde.
- Dada la alta variación de los valores la estimación vía muestra es cara.
- Crecimientos difíciles de medir.

Mediciones en base volumen

Ventajas

- Usado en la mayoría de los estudios de crecimiento y en inventarios forestales.
- La madera en pie (volumen en pie)[3] y la madera en trozas, después de la cosecha, pueden ser medidas en las mismas unidades.
- Las medidas pueden ser hechas en terreno, bajo supervisión.

---

[3] Se refiere al volumen contenido en los árboles vivos en el bosques, antes de ser cortados.

Desventajas

- El volumen no es fácil de medir directamente y es necesario estimarlo usando otras medidas (*Dap* y *H*) que tienen una alta relación con él.
- Las estimaciones de volumen pueden variar en su precisión y por lo tanto dificultar comparaciones entre rodales o bosques.
- El volumen puede no estar muy bien relacionado con el valor de la madera.