E-Book ISBN 978-987-1676-14-9.
Fecha de catalogación: 04/04/2014.

INTRODUCCION

Esta serie didáctica fue preparada en el año 1999, en oportunidad de dictarse por primera vez la asignatura "Elementos de Matemática y Estadística" del primer año de la carrera "Técnico en Viveros y Plantaciones", del plan de estudios 1999. Para nuestro equipo Cátedra, a cargo del dictado de los temas de Estadística, fue un desafío, enseñar en tan poco tiempo (poco menos de 30 horas de clase) y para alumnos que no poseían conocimientos básicos de diferenciación e integración, nociones de estimaciones por intervalo, por supuesto que pasando por un dictado previo y sintético de probabilidades y distribuciones de probabilidades. También se incluyeron los clásicos temas de la Estadística Descriptiva: tablas, gráficos y medidas de posición y dispersión.

Los resultados obtenidos pueden calificarse como positivos: el esfuerzo de nuestra Cátedra se vió recompensado por el buen rendimiento de los alumnos, los que sin duda estuvieron incentivados por el régimen promocional que tiene la asignatura. En el deseo de compartir con nuestros estudiantes esta Serie didáctica, editamos estas páginas, las que además incluyen las guías de Trabajos Prácticos utilizadas.

Cátedra de Estadística Forestal

INDICE

1.-VARIABLES,TABLAS ESTADÍSTICAS Y GRÁFICOS

2.- MEDIDAS DE POSICIÓN Y DISPERSIÓN

3.-PROBABILIDADES

CAPITULO I

Variables, tablas estadísticas y gráficos.

Estadística. Conceptos básicos

La **Estadística** es una disciplina perteneciente a la Matemática Aplicada que se dedica al estudio cuantitativo de fenómenos colectivos. Proporciona los métodos para:

- La recolección de datos
- Su ordenamiento, resumen y presentación,
- Su análisis e interpretación y
- Posterior enunciado de conclusiones.

Los cuatro pasos que se han enumerado constituyen las etapas del trabajo estadístico. En la cuarta, o sea en el enunciado de conclusiones, deben diferenciarse dos situaciones:

1. Si las conclusiones se refieren sola y exclusivamente a los datos de los que se dispone, se dice que la ***Estadística es Descriptiva.***

2. Si por el contrario, las conclusiones van más allá de los datos y se refieren a un conjunto mayor, del cual se extrajeron los datos para el análisis, se dice que la ***Estadística es Inferencial***

Las **estadísticas** (en plural) se obtienen como resultado del trabajo estadístico y están constituidas por porcentajes, promedios, tablas, gráficos y otros elementos que describen un fenómeno y ayudan a su comprensión (Ej.: estadísticas demográficas, estadísticas forestales, estadísticas del fútbol, estadísticas de accidentes de tránsito, estadísticas universitarias, etc.).

Población y muestra

Población es el conjunto de todos los individuos cuyo conocimiento interesa. La muestra es un subconjunto de la población objeto de estudio. La Estadística Inferencial trabaja exclusivamente sobre la base de muestras y extienden sus conclusiones a la Población.

Variables. Concepto y tipos.

Las variables son el objeto de estudio de la estadística. Se define a una variable como una característica capaz de asumir distintos valores o calidades. Cuando se desea estudiar alguna característica de la población se puede proceder de dos maneras:

a) Se mide u observa esa característica en cada uno de los individuos de la población, es decir se realiza un censo, el que es difícil de llevar a cabo por que insume mucho tiempo y por lo tanto mucha erogación.

b) Se mide u observa esa característica en un subconjunto de la población o muestra y luego se infieren o extienden los resultados obtenidos a la población mediante herramientas que brinda la Estadística Inferencial.

Ya se explicó que la característica objeto de estudio, que varía de un individuo a otro, es decir que puede tomar diferentes valores o cualidades se denomina **variable**. A los valores que toma esa característica se los obtiene por mediciones, conteos u observaciones que se efectúan en cada uno de los individuos que componen la muestra. Considérense los siguientes ejemplos:

Ejemplo 1: Supóngase que interesa conocer la salud de los plantines en un vivero, entonces la variable a **observar** en cada planta será el **estado sanitario**, el que podrá asumir dos valores: sano o enfermo.

Ejemplo 2: Si interesa saber el **número de semillas que germinan en cajas de Petr**i donde se ponen a germinar 6 semillas, se deberán **contar** en cada caja el nº de semillas germinadas y sus valores pueden ser: 0, 1, 2, 3, 4, 5, 6.

Ejemplo 3: Si el objetivo de un estudio fuera la **altura** alcanzada por plantas de un año de Prosopis, se debe **medir** con una vara a la variable altura la que, expresada en metros podrá tener valores entre 0 y 0.5 m.

En los tres ejemplos anteriores, el nombre de la variable y la forma de obtener sus valores está resaltado en negrita. En el primer ejemplo, los valores que puede asumir la variable son calidades, por lo que se dice que la **variable es cualitativa**. Por el contrario, en los otros dos ejemplos los valores de las variables pueden expresarse mediante números, por lo que las dos últimas **variables** son **cuantitativas**. En el caso de número de semillas germinadas, la variable toma sólo determinados valores en el intervalo que va de cero a seis por lo que se la denomina **variable cuantitativa discreta o discontinua**; cuando la variable toma los infinitos valores dentro del intervalo se dice que la **variable es cuantitativa continua**

Series de datos. Series simples

El conjunto de valores de una variable constituye una serie de datos. Se presentan a continuación series de datos referidas a los tres ejemplos que se dieron para ilustrar tipos de variables:

Ejemplo 1: Un viverista examina 12 plantines y anota su estado sanitario (S=Sano, E=Enfermo).

Generalmente las variables se representan con $x_{i,}$ de éste modo las 12 observaciones son:

x_i : S, S, E, E, E, S, S, E, S, S, S, S.

El subíndice " i " varía de 1 a 12. Así x_1 = S; x_2 = S; x_3 = E; . . . x_{12} =S.

Ejemplo 2: Un técnico examina 30 cajas de Petri en las que se colocaron para germinar seis semillas y cuenta el nº de semillas germinadas en cada una de ellas. Los valores de las 30 observaciones son los siguientes:

x_i : 4, 1, 6, 2, 4, 2, 4, 2, 4, 6, 3, 5, 3, 2, 5, 4, 0, 5, 4, 2, 4, 5, 3, 5, 3, 5,4, 3, 6, 2.

El subíndice "i" va desde 1 a 30 y entonces x_1 = 4; x_2 = 1; x_3 = 6; . .; x_{30} =2.

Ejemplo 3: Un viverista mide la altura alcanzada por 25 plantas de Prosopis de un año de edad, obteniendo los siguientes valores:

x_i (cm): 38, 14, 44, 11, 9, 21, 39, 28, 41, 4, 35, 24, 36, 12, 20, 31, 24, 25, 10, 21, 11, 36, 37, 20, 26.

Ahora "i" va desde 1 a 25, entonces x_1 = 38; x_2 = 14; x_3 = 44; . . .; x_{25} =26.

Los datos en bruto, tal cual fueron obtenidos, sin agrupar constituyen una serie simple.

Tablas y gráficos

Organización de datos categóricos o cualitativos.

Cuando la masa de datos obtenidos es muy grande y éstos están desordenados, no dan información alguna. Conviene por lo tanto ordenarlos y tabularlos, haciendo uso de tablas estadísticas, que deben confeccionarse de tal modo que los datos resulten fáciles de ser leídos e interpretados. Con los datos del ejemplo 1 se puede construir una tabla de frecuencias. Una tabla de frecuencias para variables cualitativa, es una tabla que asocia cada categoría de la variable con el número de veces que se repite la categoría.

Tabla 1. Estado sanitario de 12 plantines de un vivero

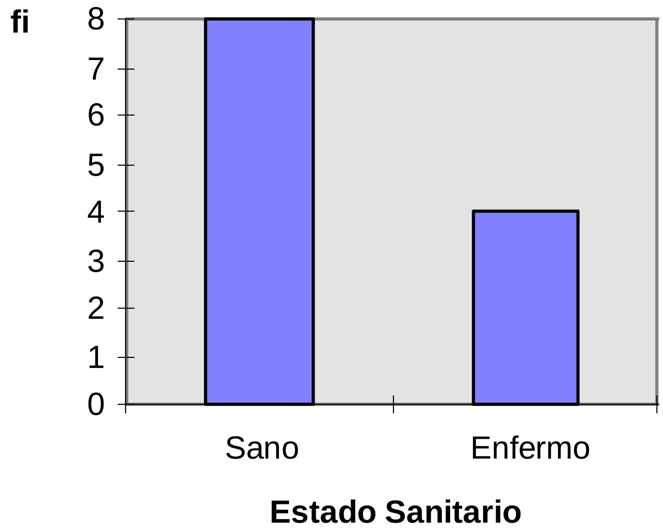
i	**Categorías:x_i** (Estado sanitario)	**Frecuencias: f_i** (nº de plantas)	**Porcentajes**: %
1	Sano	8	67
2	Enfermo	4	33
	Total	12	100

Fuente: Datos ficticios

El nº de veces que se repite cada categoría de la variable se denomina ***frecuencia absoluta*** y se la simboliza con f_i. La suma de las frecuencias absolutas, es igual al nº total de observaciones, en éste caso 12 ($\sum_{i=1}^{2} f_i$ =12). Nótese que " i " ahora se refiere a las categorías, x_1 = Sano, f_1 = 8; x_2 = Enfermo, f_2= 4.

La tabla de frecuencias, es la más sencilla de las tablas y es una tabla de simple entrada pues los individuos se clasifican según una única variable, estado sanitario en el ejemplo.

Los datos organizados en **tabla de simple entrada para variable cualitativa**, pueden presentarse mediante gráficos, que tiene la finalidad de que la información entre por los ojos. El gráfico que puede usarse en éste caso es el **gráfico de barras.**

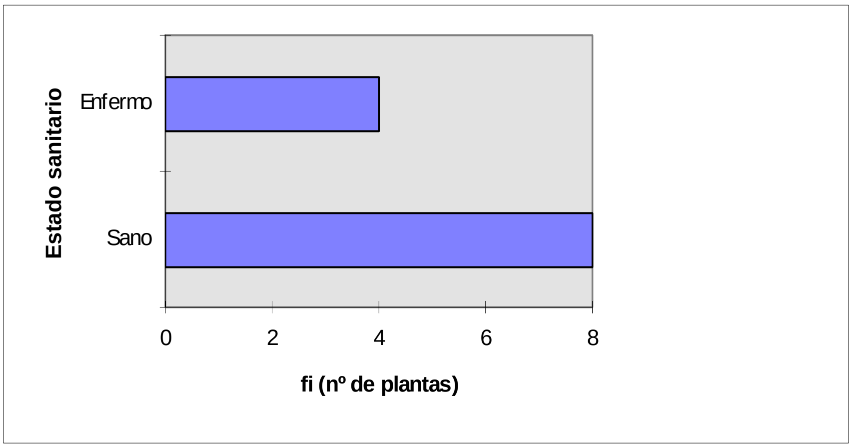


Fuente: Datos ficticios

Gráfico 1a. 12 plantines de un vivero según estado sanitario.

Para su construcción se utiliza el sistema de coordenadas ortogonales. Sobre el eje horizontal se colocan las distintas categorías de la variable en estudio (estado sanitario) y sobre el eje vertical con una escala adecuada, se representan las frecuencias. Se dibujan barras de ancho constante, una para cada valor de la variable, con una altura que representa el valor de la frecuencia que corresponde a cada categoría. Es conveniente que la separación entre las barras sea menor que el ancho de las mismas.

El ancho de las barras debe elegirse teniendo en cuenta el espacio disponible, el número de categorías de la variable a representar y la altura que les corresponde, con el objeto de obtener un gráfico proporcionado. Las barras pueden dibujarse en sentido vertical u horizontal. En algunos casos en lugar de rectángulos se dibuja una línea, razón por la cuál se denominan gráfico de líneas.



Fuente: datos ficticios

Gráfico 1b. 12 plantines de un vivero según estado sanitario

En algunos trabajos es necesario calcular frecuencias relativas. La **frecuencia relativa** de una categoría es la proporción de veces que ocurre dicha categoría. Se obtiene dividiendo la frecuencia absoluta de cada categoría entre la suma de las frecuencias de todas las categorías. La suma en éste caso es $f_1 + f_2 = 4 + 8 = 12$, y se expresa literalmente mediante el signo $\sum$ que se denomina sumatoria, así

$$\sum_{i=1}^{i=2} fi = f_1 + f_2 = 4 + 8 = 12$$

a la frecuencia relativa de la clase iésima se la simboliza con fr_i y se la calcula de la siguiente manera:

$$fr_i = \frac{f_i}{\sum f_i}$$

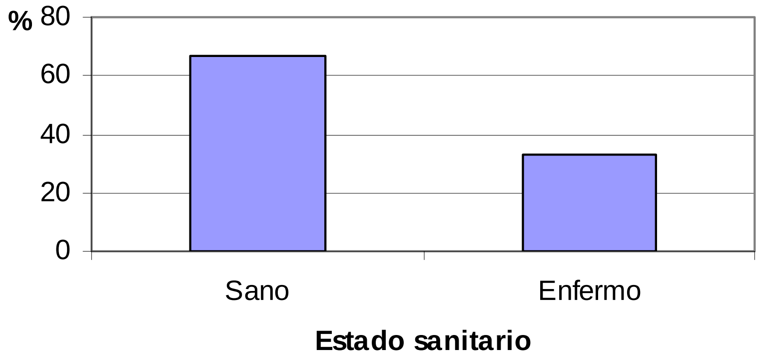
La suma de las frecuencias relativas es siempre igual a 1. Si se multiplica las frecuencias relativas por 100,se obtienen **porcentajes**. En éste ejemplo sería:

Tabla 2. Estado sanitario de 12 plantines de un vivero

i	x_i (Es.sanitario)	f_i	fr_i	**Porcentajes:** %
1	Sano	8	8/12=0.67	67
2	Enfermo	4	4/12=0.33	33
	Total	12	1.00	100

Fuente: Datos ficticios

Se pueden representar los datos de la tabla 2 mediante un gráfico de barras, sólo que en el eje vertical van los porcentajes.



Fuente: Datos ficticios

Gráfico 2. Plantines de un vivero (en %) según estado sanitario.

Otro gráfico adecuado para representar series de frecuencias de variable cualitativa es **el gráfico de sectores circulares**, llamado gráfico de tortas o pie charts. Éste no utiliza el sistema de coordenadas cartesianas para su representación. Se elige un radio y se construye un círculo que representará el total de frecuencias.

Tabla 3. Plantas producidas en el año 1999 en el vivero del INSIMA

Especies	**f_i (nº de plantas producidas)**
Grevillea	2000
Jacarandá	2000
Algarrobo	3500
Casuarinas	1200
Total	8700

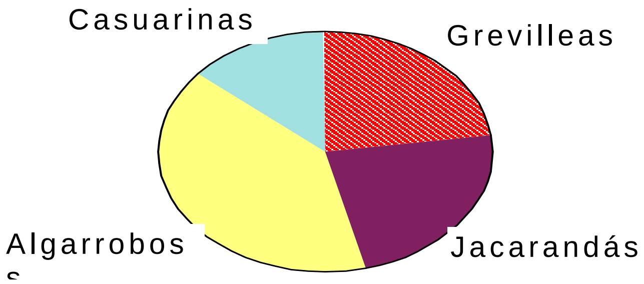
Fuente: INSIMA

Con un círculo de 3 cm de radio(el valor del radio se elige según el espacio que se disponga para el gráfico) se representa el total de plantas producidas (8700) al que, en consecuencia, le corresponde un ángulo de 360°. Mediante regla de tres se calculan los grados correspondientes a los sectores que representarán las distintas categorías de la variable especie.

Si el total de 8700 se representa con 360º
las grevilleas que son 2000 se representarán con

$$x = \frac{360 \times 2000}{8700} = 82.76º$$

De la misma manera se calcula para cada una de las especies restantes. Los valores son: 82.76º para jacarandá, 144.83º para algarrobo y 49.65º para Casuarinas. La suma de dicha columna debe ser igual a 360º.



Fuente: INSIMA

Gráfico 3: Plantas producidas en el año 1999 en el vivero del INSIMA, según especies

Si se desea representar la cantidad de plantines producidos durante dos años, por ejemplo, en vez del gráfico de barras simples, se puede usar el **gráfico de superficies**. Éste gráfico sirve para representar magnitudes por medio de superficies, de tal manera que la proporción entre las superficies sea la misma que la que existe entre las magnitudes que ellas representan. Se tiene la siguiente tabla y se quiere representar la producción de los dos años 1998 y 1999 mediante círculos.

Tabla 4. Plantas producidas en el vivero del INSIMA en los dos últimos años

Año	**Nº de plantas**
1998	5000
1999	8700

Fuente: Vivero INSIMA

Para respetar el principio de proporcionalidad básico en el gráfico de superficies se debe cumplir la siguiente relación:

$$\frac{5000}{8700} = \frac{S_{98}}{S_{99}}$$

Donde S_{98} y S_{99} corresponden a las áreas de las figuras que representan a 5000 y 8700 plantas respectivamente. Las figuras geométricas usadas son triángulos, rectángulos, cuadrados o círculos.

Si se utiliza el círculo para representar las superficies los pasos a seguir son:

1- Se elige un valor del radio (depende del espacio disponible para realizar el gráfico), que corresponde al mayor total a representar. Por ejemplo se elige un radio de 3 cm para dibujar el círculo cuya superficie representará la producción de 1999 o sea 8700 plantas.

2- Para poder dibujar proporcionalmente un círculo que corresponda al año 1998, es decir cuya superficie represente 5000 plantas se procede como sigue. Se calcula la superficie que corresponde al año 1999.

$$S_{99} = \pi \times r^2 = 3..1416 \times 3^2 = 28.2744 cm^2$$

La superficie correspondiente al año 1998, para que se mantenga la proporcionalidad, es

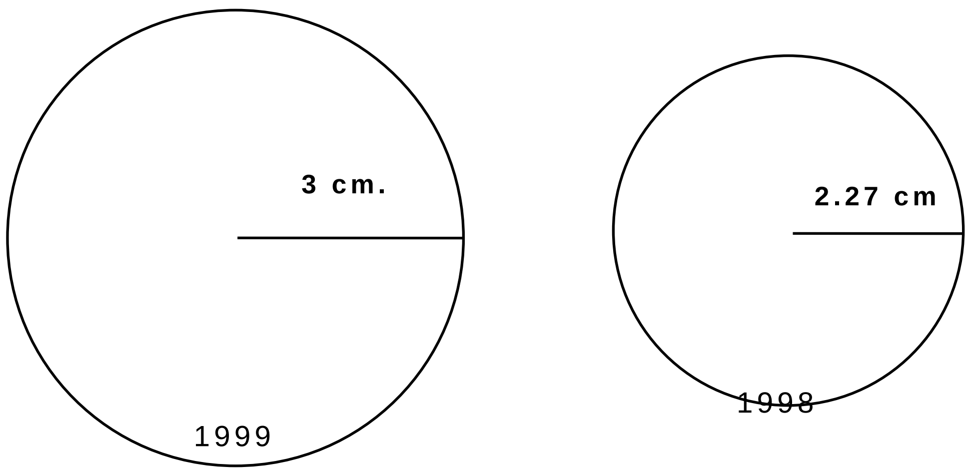
$$S_{98} = \frac{5000}{8700} \times 28.2744 = 16.2497 cm^2$$

3. - Ahora, se debe calcular el radio del círculo cuya superficie es 16.2497 cm^2.

Se sabe que $S_{98} = 16.2497 = \pi \times r^2 \quad \Rightarrow r = \sqrt{\frac{16.2497}{\pi}} = 2.27 cm$

O sea que la superficie de un círculo de radio 2.27cm representará la producción de plantas de 1998 y cumplirá con el principio de proporcionalidad:

$$\frac{5000}{8700} = \frac{16.2497}{28.2744}$$



Fuente: Vivero INSIMA

Gráfico 4. Producción de plantas en el INSIMA durante 1998 y 1999.

Se pueden combinar los gráficos de superficies y sectores tal como se muestra más adelante (Tabla 12 y Gráfico 12).

Variables cuantitativas

Para el caso de **variables cuantitativas discretas**, la tabla de frecuencias se construye de la siguiente manera: se ubica el valor mayor y el menor valor de la variable (en el ejemplo 2 del n° de semillas germinadas en un grupo de seis semillas, el menor valor es cero y el valor mayor 6), se colocan todos los valores correspondientes en la primera columna de la tabla, y luego se ve cuántas veces están repetidos dichos valores. La tabla resultante es:

Tabla 5. Cajas de Petri clasificadas según el número de semillas germinadas.

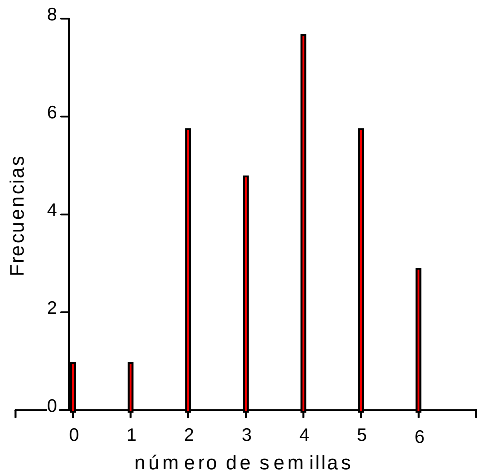
X_i	f_i
0	1
1	1
2	6
3	5
4	8
5	6
6	3

Fuente datos ficticios

La diferencia que existe entre cada clase es constante e igual a 1.

La tabla de frecuencias para variables cuantitativas discretas se representa mediante un gráfico de barras simples o, cuando el ancho de las barras es una línea recibe el nombre de barras lineales o gráfico de bastones. En la abscisa van los valores de la variable y se levanta para cada uno de ellos una línea de altura igual a la frecuencia.

Gráfico 5: Cajas de Petri según el número de semillas germinadas.



Para el caso de **variables cuantitativas continuas** como los datos del ejemplo 3 (altura de plantas de Prosopis de 1 año) que fueron obtenidos por medición, se recomienda construir intervalos de clase, cuya amplitud depende de la cantidad de intervalos que se deseen construir y la cantidad de datos que posee la serie simple. Es recomendable que los intervalos de clases sean iguales, es decir que la amplitud de los mismos (a) sea constante. La técnica a emplear para el agrupamiento de una serie simple de variable cuantitativa continua es sencilla. Se transcribe la serie.

Xi (cm): 38, 14, 44, 11, 9, 21, 39, 28, 41, 4, 35, 24, 36, 12, 20, 31, 24, 25, 10, 21, 11, 36, 37, 20, 26.

1. -Se ubica el valor mayor que toma la variable (44 cm) y el valor menor (4 cm).

2. - Se obtiene la diferencia, la que se denomina Rango o amplitud de variación y se designa con la letra R.

$$R = x_{max} - x_{min} = 44 - 4 = 40$$

3. –El número de intervalos se puede calcular con la siguiente fórmula:

$$n° \text{ de intervalos} = \frac{\log(n+1)}{\log(2)}$$

dónde n: n° de valores de la serie o tamaño de la muestra
log: logaritmo decimal

$$\text{n°de int erv.} = \frac{\log(25+1)}{\log(2)} = 4.7004 \approx 5 \text{int ervalos}$$

Cuando en la variable que se estudia existen intervalos predeterminados, como en el caso de los diámetros de los árboles, el número de clases o intervalos dependerá de la amplitud que se usa habitualmente.

4. - El rango se divide entre el nº de clases o intervalos de clases 5 para éste ejemplo, (se recomienda que no sea menor que 5, ni mayor de 15) obteniéndose una idea aproximada de la longitud o amplitud del intervalo de clase.

$$a = \frac{Rango}{n^{o}\, de \text{int} ervalos} = \frac{40}{5} = 8$$

Éste valor de amplitud es orientativo, por lo que se decide tomar una amplitud de intervalo 10 cm para facilitar el agrupamiento.

5.- Se delimitan las clases buscando preferentemente valores enteros para sus límites. Se debe elegir el límite inferior del 1^{er} intervalo de tal manera que contenga al menor valor de la serie (4 cm). La elección recae en el 0. El límite superior del 1^{er} intervalo, se obtiene sumando al Li del 1^{er} intervalo la amplitud.

Li del 1er intervalo = 0
Ls del 1er intervalo = Li + a= 0 + 10 = 10

El límite inferior del 2do intervalo debe coincidir con el límite superior del primer intervalo.

Li del 2do intervalo = 10
Ls del 2do intervalo Li + a= 10 + 10 = 20

El límite inferior del 3er intervalo debe coincidir con el límite superior del 2do intervalo, y así sucesivamente, hasta que el límite superior del último intervalo, contenga el valor observado más alto de la variable.

6.- Una vez formadas las clases se procede al conteo, que consiste en determinar el nº de observaciones (frecuencias) de cada clase. Una manera sencilla de hacerlo es leyendo la serie simple y ubicando mediante marcas cada valor de la variable en su clase correspondiente. De ésta manera cuando se termine de pasar lista a la serie simple, el agrupamiento ha sido efectuado.

Tabla 6. Plantas de Prosopis de 1 año de edad, según su altura.

Intervalo de clase (altura en cm)	x_i (marca de clase)	f_i	fr_i
0 a 10	5	2	0.08
10 a 20	15	5	0.20
20 a 30	25	9	0.36
30 a 40	35	7	0.28
40 a 50	45	2	0.08
Total		25	1.00

Fuente: Datos ficticios

Un problema que se puede presentar es el siguiente: si un valor de la variable coincide con uno de los límites del intervalo, por ejemplo la altura 20 cm ¿dónde se lo ubica? ¿en el segundo o en el tercer intervalo de clase? La respuesta es: puede ubicarlo en cualquiera de los intervalos, pero si se elige un criterio se lo debe respetar hasta el final del agrupamiento. En éste ejemplo al nº 20 se lo ubica en el 3er intervalo, de la misma manera, cuando aparezca por ejemplo un valor 40, debe ser anotado como perteneciente al intervalo en el que el nº 40 se encuentra como límite inferior.

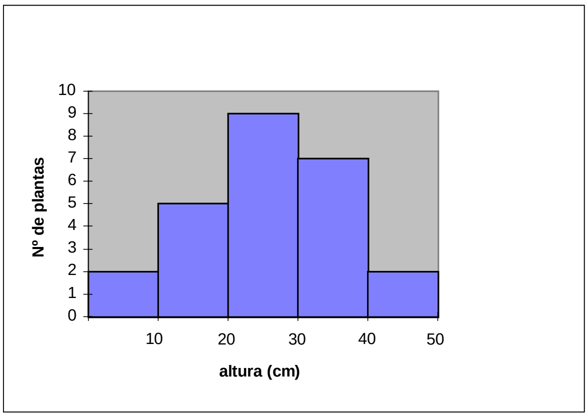
6.- Se agrega una tercera columna, titulada "marca de clase" o "punto medio de clase" que se designa con x_i que contiene los valores correspondientes a los puntos medios de cada uno de los intervalos y se calcula así.

$$x_1 = \frac{Li_1 + Ls_1}{2} = \frac{0+10}{2} = 5$$

$$x_2 = \frac{Li_2 - Ls_2}{2} = \frac{10+20}{2} = 15$$

Al efectuar el agrupamiento, se pierde detalle de la información ya que, por ejemplo, de los valores que resultaron ubicados en la primera clase, sólo se sabe ahora que se encuentran entre 0 y 10. Por eso, en caso de ser necesario asignar un valor a cada uno de ellos, como al calcular la media aritmética a partir de la tabla de frecuencias, se opta por pensar que todos tienen igual valor, que es el correspondiente al punto medio de clase.

Un gráfico adecuado para representar una serie de frecuencias de variable cuantitativa continua es **el histograma** (gráfico nº 6). Su construcción es fácil. Se utiliza el sistema de coordenadas cartesianas ortogonales. En el eje de las ordenadas (vertical) se marcan las frecuencias (f_i) y en el de las abscisas (horizontal), la variable según la cual se efectuó la clasificación (altura). Consiste en rectángulos adyacentes (uno por cada clase) con bases materializadas por la amplitud de clases (10 cm). La altura está dada por la frecuencia correspondiente a la clase. Cuando las clases son iguales, el área del histograma es proporcional a la frecuencia total.



Fuente: Datos ficticios

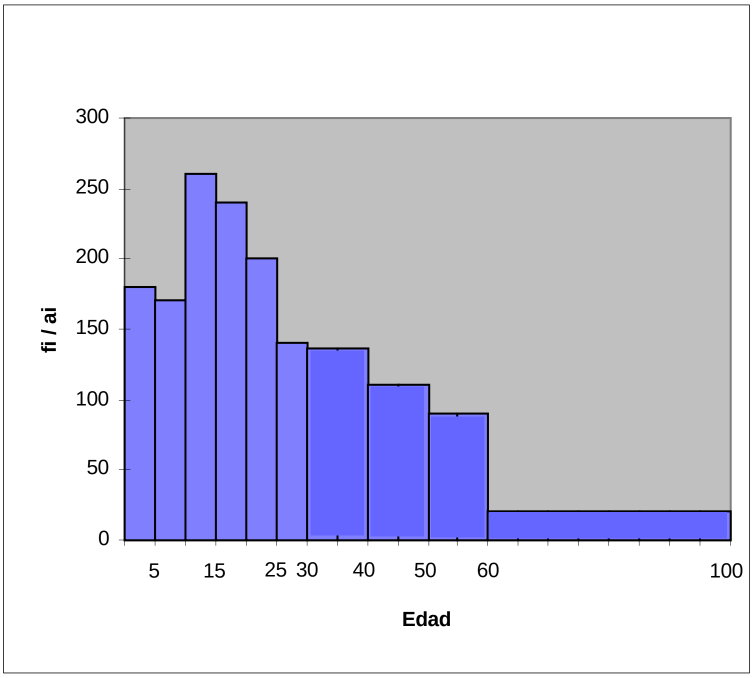
Gráfico 6.Plantas de Prosopis de un año de edad según su altura.

Pero, muchas veces y por diversas razones, las series presentan amplitud de clase variable. Como puede observarse en la siguiente tabla:

Tabla 7.Distribución de edades de una población

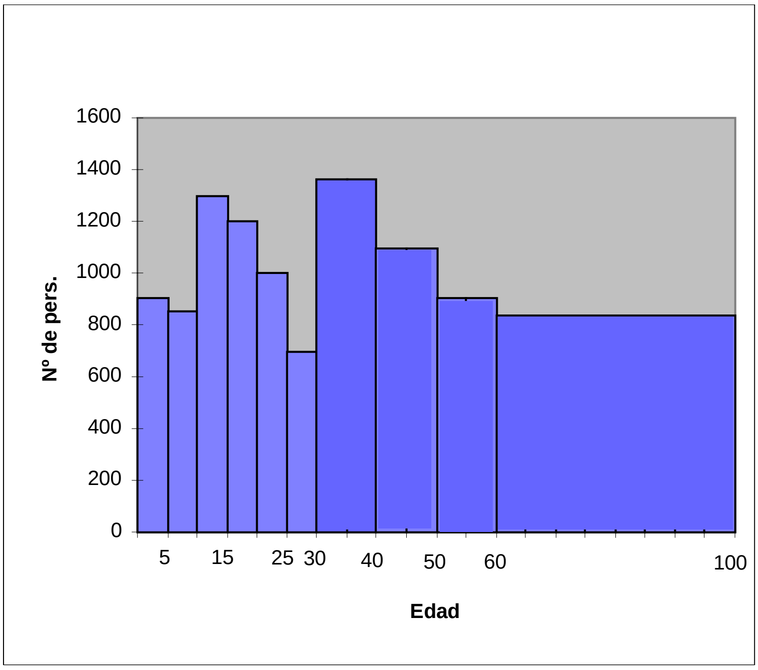
Edades	Nº de personas
0 a 5	900
5 a 10	850
10 a 15	1.300
15 a 20	1.200
20 a 25	1.000
25 a 30	700
30 a 40	1.360
40 a 50	1.100
50 a 60	900
60 a 100	840

Fuente Datos ficticios



Fuente datos ficticios

Gráfico 7a. Distribución de la población según edades



Fuente: Datos ficticios

Gráfico 7b .Distribución de la población según edades

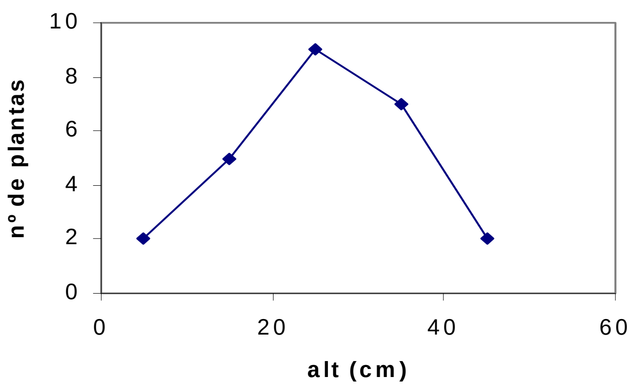
Comparando ambas representaciones gráficas, se nota claramente que la información aparece falseada en el gráfico 7b, pues en ella se ve que hay más personas comprendidas entre 60 y 100 años cuando los datos no expresan lo mismo.

Para que la representación gráfica sea correcta y las frecuencias de las distintas clases comparables, es necesario expresar las frecuencias teniendo en cuenta la amplitud de clase a la cual pertenecen, para ello se divide la frecuencia entre la amplitud de la clase.

Otro gráfico adecuado para representar la serie de frecuencias de variable cuantitativa continua es el polígono de frecuencias (gráfico 8). Se emplea para su realización el sistema de coordenadas cartesianas ortogonales. Se coloca la variable clasificadora en el eje horizontal y las frecuencias en el vertical.

La construcción es sencilla, se marcan tantos puntos como pares de valores (xi,fi) o sea marcas de clase, frecuencias haya en la tabla. En la tabla Nº 6 vemos que hay 5 pares de valores; el primer par tiene abscisa 5 y ordenada 2 y así sucesivamente hasta marcar el quinto par. Luego se unen los puntos mediante trazos rectos. Algunos autores, en su afán de mantener la proporcionalidad entre la superficie y la frecuencia aconsejan cerrar el polígono de frecuencias uniendo el primer punto con la marca de clase inmediata anterior y el último punto con la inmediata superior; en éstos dos casos la unión de los puntos se realiza con trazos cortados.

La principal ventaja de los polígonos de frecuencias consiste en que ellos permiten dibujar en el mismo sistema de eje dos o más polígonos correspondientes a series diferentes que tengan similar posición sobre el eje de las x, así se puede compararlos, lo cual resulta engorroso efectuar con los histogramas a causa de la superposición de las superficies de de los rectángulos.



Fuente: Datos ficticios

Gráfico 8.Plantas de Prosopis de un año de edad según su altura.

Con los ejemplos anteriores se ha representado gráficamente y ordenado datos relativos a una variable de la población, tal como la altura, o el estado sanitario de las plantas de un vivero. Se vió que, cuando el número de valores obtenidos en una distribución es pequeño, a la hora de presentarlos basta, simplemente, con enumerarlos ordenadamente, como en el siguiente ejemplo que corresponde a la nota obtenida por diez alumnos en el parcial de estadística.

X_i : 3, 3, 4, 5, 5, 7, 8, 9, 10.

Cuando el número de datos es grande, para ordenarlos se debe usar el agrupamiento en una tabla de frecuencias.

Tabla 8. Alumnos clasificados según la nota obtenida en los parciales de Estadística.

xi	1	2	3	4	5	6	7	8	9	10
fi	2	11	17	48	66	40	32	21	8	5

Fuente: Datos ficticios

Pero cada miembro de una población presenta diversos aspectos que pueden ser de interés para el técnico, y él puede necesitar clasificar a los individuos de dicha población de acuerdo a dos variables, por ejemplo le interesa medir el diámetro a la base y la altura de las plántulas del vivero. Tiene para cada individuo medido dos valores de variable. Cuando el número de individuos medidos es pequeño, se enumeran todos los pares de observaciones, si alguno de ellos aparece dos veces, se lo repite y la presentación suele hacerse de modo que una de las dos variables esté ordenada.

Tabla 9. 12 Plantas de un vivero clasificadas según el diámetro a la base y altura

Dab	8	8	9	10	11	11	12	13	14	15	15	15
Alt	119	118	121	118	120	123	119	121	119	129	127	130

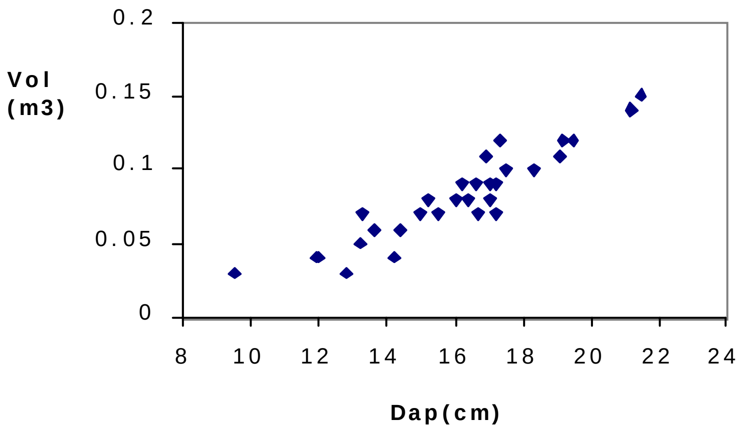
Datos ficticios

Para representar estos datos que corresponden a dos variables cuantitativas continuas se utilizan los ***gráficos de dispersión o scatter plot***, que se construye de la siguiente manera: se coloca una de las variables en las abscisas o eje horizontal, por ejemplo el diámetro y la otra variable, la altura, en el eje vertical, con sus escalas correspondientes, luego se marcan tantos puntos como pares de valores (x_i, y_i) se tengan. Se presenta otro ejemplo en la tabla 9a.

Tabla 9a. Diámetros y volúmenes de los árboles de una parcela de 576 m2 situada en una plantación de paraísos de 8 años en el Dpto. Copo (Sgo. del Estero).

Dap(cm)	Vol(m^3)	Dap (cm)	Vol(m^3)	Dap (cm)	Vol(m^3)
9.50	0.03	15.20	0.07	17.20	0.09
11.90	0.04	15.50	0.07	17.20	0.09
12.00	0.04	16.00	0.08	17.30	0.09
12.80	0.05	16.20	0.08	17.50	0.10
13.20	0.05	16.40	0.08	18.30	0.10
13.30	0.05	16.60	0.09	19.10	0.11
13.60	0.06	16.70	0.09	19.20	0.12
14.20	0.06	16.90	0.09	19.50	0.12
14.40	0.06	17.00	0.09	21.20	0.14
15.00	0.07	17.00	0.09	21.50	0.15

Fuente. Cátedra de Estadística Ftal.



Fuente. Cátedra de Estadística Ftal.

Gráfico 9. Relación diámetro (en cm) volumen (m3) de árboles de una parcela de 576 m2 ubicada en la plantación de paraíso de 9 años de edad en el Dpto. Copo (Sgo. del Estero)

Éste gráfico sirve para mostrar la relación entre las dos variables y se usa cuando para el mismo valor de xi se tiene diferentes valores de y_i. Si esto no ocurre puede utilizarse el gráfico lineal, que se construye de igual manera que el anterior, con la única diferencia que se unen los puntos. Éste gráfico, se suele emplear, especialmente, en los casos donde la variable que se representa en el eje horizontal es el tiempo. De éste modo se puede ver la evolución de la otra variable en el período considerado. Pueden representar simultáneamente en el mismo gráfico dos o más variables, como se observará al representar gráficamente los datos de tabla Nº 10

Tabla Nº 10. Temperatura del suelo y del aire (ºC) registradas en el Zanjón en el mes de abril de 1999.

	Suelo	Aire		
Día	T(°C)	Tmedia °C)	Tmax(°C)	Tmin(°C)
1	20.6	16.0	18.3	14.5
2	19.5	16.1	19.7	14.1
3	19.5	16.8	19.3	15.0
4	20.1	18.8	25.6	14.7
5	20.2	18.9	29.7	10.7
6	20.6	18.3	24.6	12.7
7	20.0	17.8	24.6	13.7
8	20.0	18.3	24.8	14.4
9	20.1	18.2	25.8	13.9
10	20.3	18.4	21.0	17.0
11	20.0	18.1	22.1	14.8
12	20.0	18.7	27.3	11.3
13	20.5	18.2	27.7	9.6
14	21.9	20.6	29.5	11.9

15	20.9	15.1	19.7	11.2
16	18.3	12.6	20.6	4.3
17	17.2	9.3	22.3	-1.3
18	17.1	10.8	24.2	-0.2
19	17.6	12.9	26.1	1.3
20	18.5	14.9	27.5	4.0
21	19.4	15.4	27.9	4.9
22	20.6	19.2	30.5	9.1
23	22.7	23.4	33.4	15.8
24	24.8	25.9	34.2	21.4
25	24.6	23.2	32.9	18.3
26	22.1	17.5	23.6	13.0
27	20.4	17.6	21.1	15.1
28	20.2	16.9	27.7	7.8
29	20.7	18.1	26.2	9.8
30	20.4	20.3	24.7	17.8

Fuente: Boletín FAAI

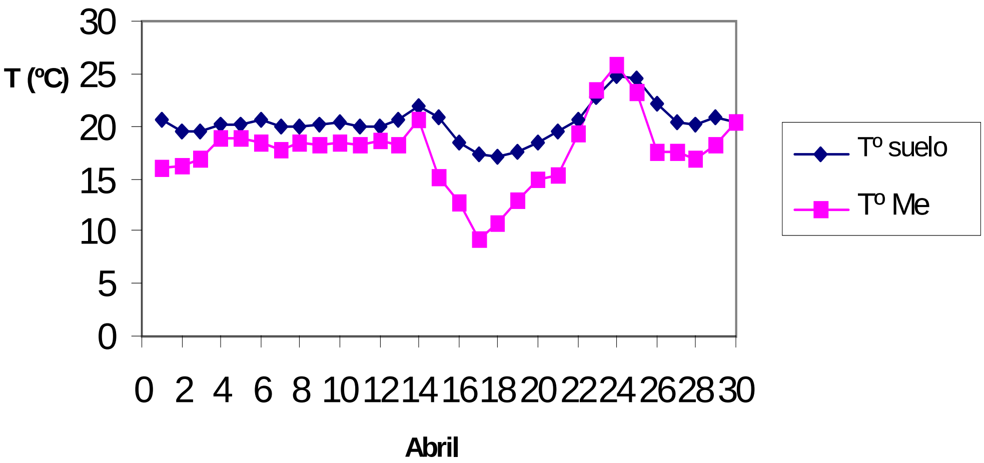


Gráfico Nº 10. Evolución de las temperaturas del Suelo (ºC) y la media del aire en El Zanjón en Abril de 1999.

Fuente: Boletín de Fac. AAI

Cuando los pares de valores son muy numerosos las tablas se presentan de la siguiente manera (tabla 11), en éste caso se dice que las tablas son de doble entrada por que los datos fueron agrupados según dos variables.

Tabla 11. Producción de plantas en un vivero según especie y tipo de envase

ESPECIE	TIPO DE ENVASE			
	TUBETES	BOLSITAS	MACETAS	TOTAL
Eucalyptus	3000	1500	500	5000
Pinus	2000	1500	100	3600
Grevilleas	1000	2000	500	2500
Algarrobo	500	2500	3000	6000
Total	6500	7500	4100	17000

Fuente: Datos ficticios

El valor de la celda se completa con la información que brinda la fila y la columna correspondiente. Por ejemplo el 3000 de la primera celda significa que en ése vivero se produjeron 3000 plantas de Eucalyptus en tubetes. Las partes de una tabla son:

- ☑ ***La matriz***, formada por la primera fila, lleva los encabezamientos de las columnas y / o la primera columna que titula a las filas,
- ☑ ***El cuerpo*** constituido por celdas.

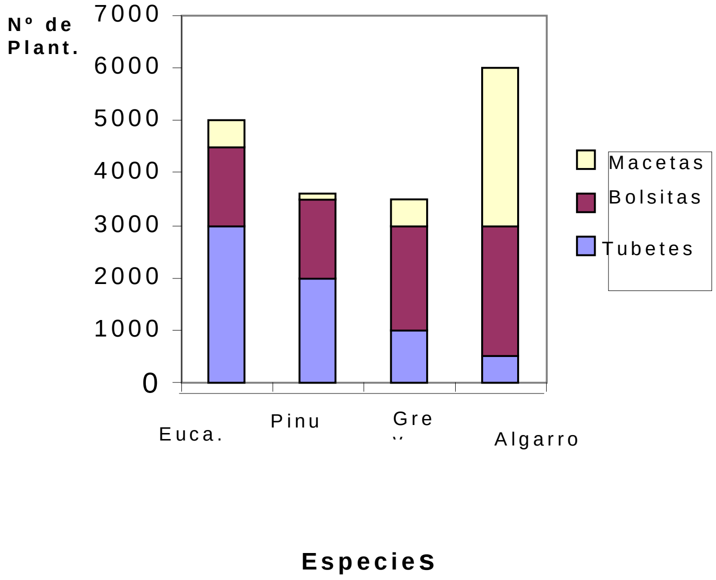
La información proporcionada por los valores de las celdas se completa con la suministrada por los encabezamientos de las filas y columnas, en las celdas se encuentra la frecuencia, es decir la cantidad de elementos o individuos que poseen las dos características.

El gráfico que se utiliza sirve para representar éste tipo de tablas es el ***gráfico de barras compuestas*** (gráfico 11a) y el ***gráfico de barras agrupadas*** (gráfico 11b).

En la tabla 11 las variables clasificadoras son especie (variable cualitativa) y tipo de envase (variable cualitativa).

La construcción del gráfico de barras compuestas es sencilla. Se comienza dibujando las barras como si fueran simples es decir con las alturas correspondientes a los totales y luego se yuxtaponen los valores parciales hasta alcanzar el de su suma. En el ejemplo, para Eucalyptus, se procede de la siguiente manera: se marca una barra de altura 5000, en ella se indica la primera subdivisión que puede ser tubetes con el valor 3000. Para bolsitas se aconseja proceder a la suma de tubetes + bolsitas= 3000 + 1500 = 4500. Se marca la segunda división correspondiente a bolsitas: la porción comprendida entre 3000 y 4500, lo que resta de la barra corresponde a producción en macetas.

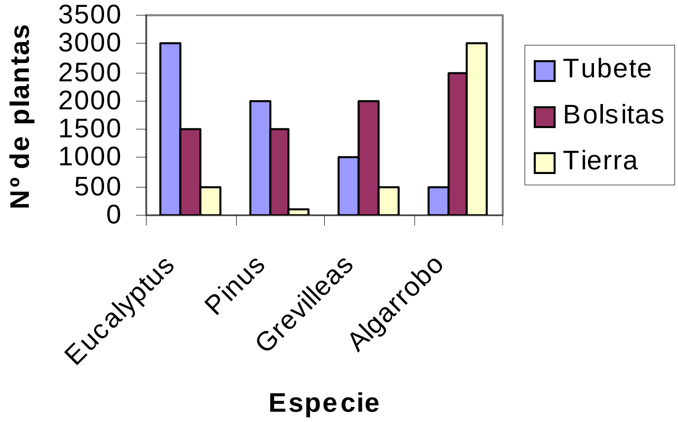
Gráfico de barras agrupadas



Fuente: Datos ficticios

Gráfico 11a. Producción de plantas de un vivero, según especie y tipo de envase

Sirven para representar fenómenos similares a los que originan barras compuestas. La diferencia con éstas estriba en que, para cada valor de la variable independiente “x” en éste ejemplo especies, se dibujan grupo de barras. El número de barras en cada grupo es el del número de categorías de la segunda variable.



Fuente:datos ficticios

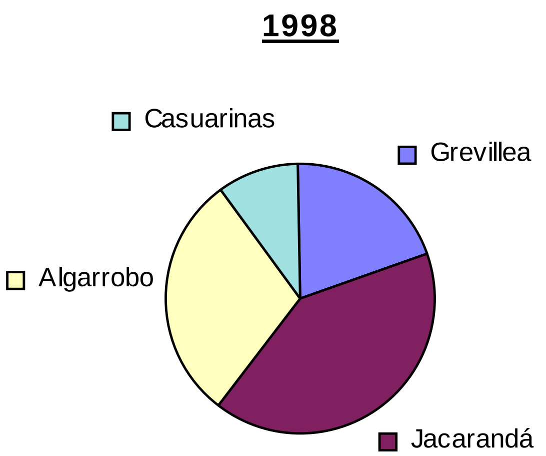
Gráfico 11b. Producción de plantas de un vivero, según especie y tipo de envase.

Otro gráfico que se puede utilizar es el gráfico de superficies combinado con el de sectores circulares. Las superficies se utilizan para representar los totales de producción y se discrimina las distintas especies mediante sectores.

Tabla12. Producción de plantas en el INSIMA en los años 1998 y 1999, discriminada por especies

Especie	Nº de plantas producidas en 1998	Nº de plantas producidas en 1999
Grevillea	1000	2000
Jacarandá	2000	2000
Algarrobo	1500	3500
Casuarinas	500	1200
Total	5000	8700

Fuente: INSIMA



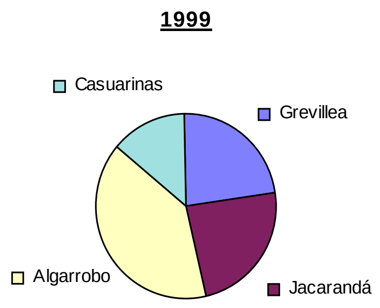


Gráfico 12. Producción de plantas en el vivero del INSIMA durante 1998 y 1999, según especies.

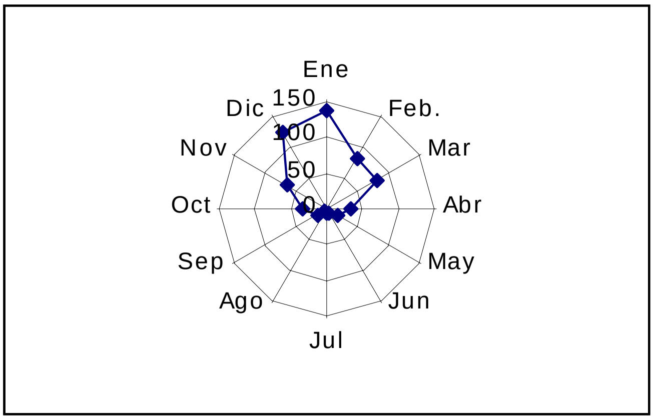
Otros tipo de gráficos que se observan en trabajos científicos y revistas son los llamados ***gráficos en espiral*** (gráfico 13). Se lo llama también gráfico de coordenadas polares. Sirven para representar la relación entre dos variables cuantitativas, especialmente cuando la independiente es cronológica y a intervalos iguales. También suele utilizarse cuando "x" indica dirección, por ejemplo procedencia de los vientos.

El valor de "x" está dado por el ángulo y el de "y" por la distancia desde el origen, sobre la línea que marca éste ángulo. Son gráficos muy efectivos para mostrar fenómenos inflacionarios, configurando en estos casos, una verdadera espiral que es la que da origen a su nombre.

Tabla Nº13.Precipitaciones medias mensuales en Sgo. del Estero en el período 1981-90

Mes	Ene	Feb.	Mar.	Abr.	May	Jun.	Jul.	Ago.	Set.	Oct.	Nov.	Dic.
Pp mm	136.3	80.8	78.2	33.5	18.3	6.6	5.6	2.4	13.6	34.3	63.7	120.4

Fuente: Ing. Pedro Boletta



Fuente:Ing. Pedro Boletta CátedraClimatología Forestal

Gráfico 12. Precipitación en valores medios (en mm) para Santiago del Estero, correspondiente al período 1981-1990.

Otro tipo de gráficos son los ***gráficos de figuras o pictogramas***. Son los más indicados para publicaciones de divulgación popular, por su fácil e inmediata interpretación. Consisten en dibujos esquemáticos y relacionados con el fenómeno a representar. Cada figura es equivalente a una cantidad determinada, preferentemente entera, de unidades de la variable dependiente y el número de unidades no su tamaño, es proporcional a la magnitud a representar.

Cartogramas: Se emplean cuando es importante señalar la distribución geográfica de un determinado acontecimiento, razón por la cual se construyen sobre planos o mapas.

Cartogramas de señalización: Sirven para indicar la distribución de una variable cualitativa sobre una base geográfica. Mediante figuras, colores o diferentes rayados se señala que hay en lugares determinados.

Cartogramas de densidad: además de indicar que hay y dónde, de ellos se puede obtener la información de cuánto hay. Mediante diferente rayado o colores y también utilizando barras o gráficos de superficies sobre la base geográfica, se puede expresar la cuantía del fenómeno como así también su ubicación. Suelen utilizarse pictogramas, gráficos de líneas, en general cualquiera de los descriptos, sobre el mapa o plano.

Resumiendo los datos se ordenan, clasifican y presentan en formas de tablas. Las **tablas** pueden de ser de simple entrada(cuando los individuos se clasifican según una variable), de doble entrada(cuando los individuos se clasifican según dos características)y de triple o más entradas (cuando se clasifican los datos según tres variables o más variables).Las tablas se complican a medida que se agregan más variables, por lo tanto es preferible varias tablas sencillas a una complicada.

Toda tabla debe llevar título, el cuál debe responder a las preguntas ¿Según?, ¿Qué?, ¿Cuándo? y ¿Dónde?.

No se debe olvidar la fuente de datos que indica de donde proviene la información.

Se debe incluir los totales

En caso de expresar los datos en porcentajes, deben indicarse los totales de los cuales provienen.

Con respecto a los **gráficos**, éstos constituyen una de las formas más útiles de presentación de datos estadísticos. Su importancia reside en las múltiples formas que pueden adoptar, lo que permite su aplicación a una amplia gama de finalidades: didácticas, de investigación, etc. Sirven para mostrar la relación entre una o más variables. La variedad de tipo de representaciones gráficas exige una cautelosa elección de acuerdo a su finalidad. La selección de la presentación gráfica debe, por lo tanto tener los siguientes aspectos:

Tipo de análisis estadístico;características y número de los fenómenos o variables a representar y público al que va dirigido.

Recomendaciones para la construcción correcta de un gráfico.

Una vez elegido el tipo de gráfico adecuado, es conveniente no descuidar las siguientes consideraciones:

*Decidir cuál de las variables es la independiente “x” y cuál la dependiente “y”.
*La representación gráfica debe ser sencilla, simple y explicarse por sí misma.
*Título se coloca encabezando el gráfico y debe responder a las preguntas; qué, según, cuándo, dónde?.
*Fuente de datos. Se coloca al pie del gráfico.
*Escalas se elige de tal modo que no alteren la objetividad de la representación, hecho éste muy utilizado para fines publicitarios donde es común ver escalas construidas con el propósito de alterar el fenómeno exagerando ventajas y enmascarando la realidad, o lo que es peor aún eliminando la graduación de los ejes, evitando de ésta forma todo patrón de comparación. Las escalas deben construirse buscando obtener como resultado un dibujo armónico y proporcionado.
*Debe nominarse los ejes de modo tal que no quede duda alguna

acerca de las variables que en ellos se representan.
*No olvidar el corte de ejes en caso de ser necesario. Éste debe efectuarse entre el 0 y el valor mínimo a representar.
*Aclaración de las unidades de representación
*Las referencias serán colocadas al pie o al costado del gráfico.
*En caso de usarse abreviaturas, éstas serán aclaradas con la debida extensión, en el renglón siguiente al correspondiente a las fuentes.
*En lo posible acompañar los gráficos con las tablas estadísticas que lo originen.
*Si el trabajo lo requiere y es necesario expresar algunos valores en %, deben consignarse las cifras de las cuales provienen éstos porcientos.

Clasificación

A. Gráficos con coordenadas.

A.1.Coordenadas ortogonales.

1) Histogramas
2) Polígonos de frecuencias
3) Barras simples, compuestas, agrupadas.
4) Lineales
5) De siluetas
6) De fajas.

A.2. Coordenadas pseudoortogonales.
A.3 Coordenadas no ortogonales.
1.Polares
2.Triangular equiláteras

B.Gráfica sin coordenadas

B.1. De figuras o pictogramas.
B.2. De superficies: simples (triangulares, cuadrangulares, rectangulares, etc.) y compuestos (triangulares, cuadrangulares, rectangulares, sectores circulares, etc.
B.3. Cartogramas: 1) de señalización y 2) de densidad
B.4. De volúmenes: simples y compuestos. (Piramidales, cúbicos, prismáticos, cilíndricos,etc.

CAPITULO II.

MEDIDAS DE POSICIÓN Y DISPERSIÓN

INTRODUCCIÓN

En todo trabajo estadístico luego de recolectar los datos, ordenarlos y agruparlos en tablas y presentarlos gráficamente, es preciso extraer alguna información que permita describir la población de la cual se extrajeron los mismos.

Existen algunas medidas que resumen los datos, es decir que nos permiten representarlos con un único valor; éstas medidas pueden proporcionar información referida a la posición del conjunto de datos en el eje de las x y se llaman **Medidas de Posición** y otras que miden como se distribuyen los datos alrededor del valor central y que se denominan **Medidas de Dispersión.**

Cuando las medidas de posición nos indican además el centro del conjunto de datos, se denominan **Medidas de Tendencia Central.** Hay otras medidas indican únicamente localización o ubicación de determinados valores en la serie son los: cuartiles, deciles y percentiles y se denominan **medidas de localización**.

MEDIDAS DE TENDENCIA CENTRAL

Según el criterio usado para determinar el centro del conjunto de datos se distinguen las siguientes medidas : media aritmética, mediana, modo y media cuadrática.

- **MEDIA ARITMÉTICA**

a) <u>Cálculo de la media aritmética en series simples</u>

Es quizás la más conocida y usada, se la llama también promedio; se la obtiene al dividir la suma de todos los valores de la serie entre la cantidad valores sumados. Se representa con $\bar{x}$, y considerando una serie simple con **n** observaciones se calcula de la siguiente manera

$$\bar{x} = \frac{\sum x_i}{n}$$

Ejemplo:

Se dispone de las siguientes alturas de plantas en cm. y se quiere averiguar cual es la altura promedio:

x_i = altura de plantas en cm.

x_i = 15; 16; 12; 14; 11

$$\bar{x} = \frac{15+16+12+14+11}{5} = \frac{68}{5} = 13{,}6cm$$

PROPIEDADES DE LA MEDIA ARITMETICA

- **Es reproductora del total**.

Esta propiedad permite conocer totales.

Ejemplo:
Sí en una plantación de paraísos a los 9 años de edad, en el Departamento Alberdi, Pcia. De Santiago del Estero, el volumen promedio por ha es de 44.17 m^3, ¿ cuál es el volumen en la superficie total que es de 125 has?. ?
Volumen total = 44.17 m^3 / ha. * 125 has. = 5521.25 m^3.

- **La suma de los desvíos con respecto a la media aritmética es siempre igual a cero.**

En el ejemplo de las cinco alturas de plantas el promedio era igual a 15 cm. $\left(\bar{x} = 15\right)$

Alturas (x_i)	$d_i = x_i - \bar{x}$
15	1.4
16	2.4
12	-1.6
14	0.4
11	-2.6
	$\sum(d_i)=0$

- **Es muy sensible a valores extremos**.

Si por equivocación al pasar los datos en el ejemplo de las cinco plantas colocamos 56 en vez de 16 cm, la media toma el valor 21,6 cm por lo que deja de representar el centro del conjunto de datos, alejándose hacia el valor extremo.

- **La media aritmética ocupa el lugar correspondiente al centro de gravedad y constituye el punto de equilibrio de los datos.**

- **La suma de los cuadrados de los desvíos con respecto a la media aritmética es mínima.**

$$\sum d_i^2 = \sum \left(x_i - \bar{x}\right)^2 = mín$$

En el ejemplo que se venía desarrollando, si se eleva al cuadrado los desvíos con respecto a la media y se los suma se tiene:

$1.4^2 + 2.4^2 + (-1.6)^2 + 0.4^2 + (-2.6)^2 = 17.2$

Que es el valor más bajo que se puede tener al restar cualquier valor a nuestros datos y luego elevarlos al cuadrado.
Por ejemplo, si en vez de la media restamos a nuestros datos el valor 15 y a ésta diferencias las elevamos al cuadrado se tiene

$(15-15)^2 + (16-15)^2 + (12-15)^2 + (14-15)^2 + (11-15)^2 = 27$

Se comprueba de esta manera la propiedad citada anteriormente pues 17.2 es menor que 27.

b) **Cálculo de loa media aritmética en series de frecuencias**

Como en una serie de frecuencias, **f_i** nos indican las veces que se repite el valor de la variable, debemos considerarlas en el cálculo de la media aritmética. Deseamos obtener la altura media d las plántulas de un vivero, los datos se presentan en la Tabla 14.

Tabla N° 14.Altura de plantas (en cm.) de un vivero

x_i	f_i
11	10
12	5
14	8
15	7
23	2
Total	32

FUENTE:Datos ficticios

donde

x_i: altura de plantas en cm.

f_i : número de plantas que poseen esas alturas

$\bar{x} = (11+11+...+11+12+...+12+14+...+14+15+...+15+23+23)/32$

Esto se podría calcular de la siguiente manera

$$\bar{x} = \frac{11*10 + 12*5 + 14*8 + 15*7 + 23*2}{32} = 13.53\text{cm.}$$

ahora expresando literalmente la fórmula de la media aritmética es:

$$\bar{x} = \frac{\sum x_i * f_i}{\sum f_i}$$

Considere ahora el cálculo de media aritmética del diámetro de ejemplares de álamos de una parcela, en una plantación de Santiago del Estero. Los datos figuran en la tabla 15

Tabla N°15. Ejemplares de álamos de una parcela en una plantación de Sgo. del Estero, clasificados por clases diamétricas.

Clases de diámetro en cm	x_i	f_i	x_i * f_i
4.3 - 5.3	4.8	2	9.6

5.3 - 6.3	5.8	7	40.6
6.3 - 7.3	6.8	14	95.2
7.3 - 8.3	7.8	13	101.4
8.3 - 9.3	8.8	1	8.8
Total		37	255.6

Fuente: Cátedra de Estadística FCF. UNSE.

En este caso se toma el punto medio de la clase $\mathbf{x_i}$ como el valor que resume todos los que están en esa clase.

$$\bar{x} = \frac{\sum x_i * f_i}{\sum f_i} = \frac{255.6}{37} = 6.91cm$$

Cuando el promedio se obtiene con todos los datos de la población, es decir cuando se efectúa un censo, obtenemos lo que se denomina parámetro de la población y se representa y calcula de la siguiente manera, siendo N el tamaño de la población

$$\mu = \frac{\sum x_i}{N}$$

- **MEDIANA**

La media aritmética no es recomendable para representar el centro del conjunto de datos cuando en la serie existen valores extremos, pues se vio que en su cálculo intervienen todos los valores de la serie, y es sensible a ellos.

Por esta razón, en el conjunto de datos con éstas características se utiliza otra medida de tendencia central que se denomina Mediana y la representamos con “Md”. La mediana es aquel valor que divide a la serie ordenada de datos en dos partes iguales, de manera tal que a ambos lados de ella quedan igual número de valores.

Para su cálculo debemos ordenar primero los datos en forma ascendente o descendente. Si el número de observaciones es impar el valor de la mediana coincide con el valor del centro. En caso de que el número de observaciones fuera par, el valor de la mediana corresponde al promedio de los dos valores centrales.
La ubicación de ese o esos valores centrales se obtiene ubicando el o los valores que se encuentran en la posición $\frac{n+1}{2}$.

Ejemplo:
Las muestra posee tamaño impar n=5

x_i : ingresos mensuales de cinco operarios en una carpintería (en pesos)
200; 350 ; 200 ; 825 ; 150

Para calcular la mediana

1ª) Se ordena los datos:

150 ; 200 ; 200 ; 350 ; 825

2ªSe calcula la posición del valor mediano: $\frac{n+1}{2}$

$\frac{5+1}{2}$ = 3, Significa que el valor mediano es el que corresponde al 3er lugar, que en este caso corresponde a 200. Entonces

M_e = 200.

150 ; 200 ; $\boxed{200}$; 350 ; 825

Esto significa que el 50% de los operarios de esa carpintería ganan $200 o menos, o el 50% de los operarios ganan $200 o más.

La muestra posee **tamaño par** n = 6

89; 23 ; 74 ; 12 ; 46 ; 25

1ª) Se ordena los datos:

12; 23; 25 ; 46 ; 74 ; 89

2ª)Se calcula la posición del valor mediano: $\frac{n+1}{2}$

$\frac{6+1}{2} = 3.5$,

significa que está ubicada entre el 3er y 4° lugar de la serie ordenada:

12 ; 23 ; $\boxed{25}$; $\boxed{46}$; 74 ; 89

$\boxed{35.5}$ Md = $\frac{25+46}{2} = 35.5$

el valor de la Mediana se obtiene promediando los valores centrales

Para el caso de series agrupadas:

Tabla N° 16. Número de árboles atacados por insectos en una parcela

x_i	f_i	f_a
0	80	80
1	60	140
2	30	170

3	25	195
4	10	205
5	5	210
Total	211	

FUENTE: Datos ficticios

En la series de frecuencias los datos ya están ordenados, por lo que solo resta encontrar el valor central, cuya posición se encuentra ubicando el valor:

$$\frac{\sum f_i + 1}{2} = \frac{211}{2} = 105.5$$

Para ello se calculan las frecuencias acumuladas y luego ubicamos el menor valor que contiene a 105 y a 106, en éste caso coincide y es 140.

Significa que el la posición 105 y 106 tenemos el valor de variable que es 1, por lo que en éste caso M_d=1.

MODO

Es el valor de variable que más se repite. Es la única medida de posición que se puede calcular para variables cualitativas nominales, es decir en las variables cualitativas en las que no se puede establecer un orden entre sus valores.

X_I : Color de flor R: rojo N: naranja A: amarillo

X_i : A ; R ; R ; A ; N ; A ; R ; R ; R ; A ; N ; R ; R ; R

Modo : Mo : R

En la siguiente serie de frecuencias anterior, nos fijamos en la columna de frecuencias absolutas cuál es el valor más alto, en éste caso es 80, que nos indica la cantidad de árboles con ningún ataque, es decir el valor modal es cero.

MEDIA CUADRÁTICA

La Media cuadrática (Mc) es la raíz cuadrada de la media aritmética de los cuadrados de los valores de la variable.

Mc= $\sqrt{\frac{\sum x_i^2}{n}}$ en series simples

Mc = $\sqrt{\frac{\sum x_i^2 * f_i}{\sum f_i}}$ en series de frecuencias

La Media cuadrática se utiliza:

a.- Cuando se promedian valores de una variable que luego será empleada elevada al cuadrado.

b.- En oportunidades de promediar valores de variable que presenten la característica de que su suma da siempre cero. Es el caso de los desvíos con respecto a la media aritmética.

Ejemplo:

Tabla N°17. Distribución diamétrica de los árboles de un bosque irregular

DAP(cm)	3-9	9-15	15-21	21-27	27-33	33-39
f_i	62	40	31	15	12	5
x_i	6	12	18	24	30	36

Fuente: Cátedra de Estadística. FCF.UNSE

Calcular la media cuadrática.

En realidad este valor es el diámetro correspondiente a la sección normal media (DAP: diámetro a 1.30m, conocido vulgarmente como diámetro a la altura de pecho)
Aplicando la fórmula

Mc= $\sqrt{\frac{43956}{165}} = 16.32 \text{ cm}$

CUARTILES, DECILES Y PERCENTILES

Son otras Medidas de Posición que no tienen en cuenta el centro de la distribución. Se refieren a otras fracciones de la serie.
Los cuartiles son tres Q_1, Q_2, Q_3, Dividen a la serie en cuatro partes iguales. El segundo cuartil coincide con la Mediana.
Por debajo del primero quedan el 25% de los datos; por debajo del segundo el 50% de los mismos y por debajo del tercero el 75%.
Los Deciles son nueve y dividen a la serie en 10 partes iguales; los percentiles son 99 y la dividen en 100 partes iguales.

MEDIDAS DE VARIABILIDAD O DISPERSIÓN

Las Medidas de Posición no son suficientes para describir el conjunto de datos sino que es necesario tener una idea de como se distribuyen los datos alrededor del centro de la distribución. Para eso surgen las Medidas de Dispersión.

RANGO

Es llamado también amplitud total de variación de la variable. Se lo obtiene como la diferencia entre el valor máximo y mínimo de la variable.

Ejemplo:

Los siguientes son datos de temperatura (°C) durante 5 días:
x_i (^{o} C) = 22 , 26 , 27 , 26 , 34
Rango= 34 - 22 = 12

La desventaja de esta medida es que solo considera los valores extremos sin tener en cuenta el comportamiento del resto de las observaciones.
Para solucionar este problema surgen otras medidas como:

DESVÍO MEDIO

Se podría trabajar con los desvíos individuales, sumarlos y promediarlos, pero no se puede hacer esto ya que siempre su valor sería cero, por propiedad de la media aritmética.
Para solucionar el problema de signos y así poder encontrar el valor promedio podemos utilizar el valor absoluto de los desvíos.

$$DM = \frac{\sum |d_i|}{n} = \frac{\sum |x_i - \bar{x}|}{n}$$

En el ejemplo anterior la media es igual a 27

x_i	$d_i = x_i - \bar{x}$
22	-5
26	-1
27	0
26	1
34	7

$$DM = \frac{5+1+0+1+7}{5} == 2.8$$

DESVIACIÓN ESTÁNDAR

Es la media cuadrática de los desvíos.

Cuando se trabaja con muestras la desviación estándar muestral es:

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$ en series simples

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2 * f_i}{\sum f_i - 1}}$$ en series de frecuencias

Para el ejemplo de las temperaturas

$$s = \sqrt{\frac{(-5)^2 + (-1)^2 + 0^2 + 1^2 + 7^2}{4}} = 4.36$$

Para ejemplificar el caso de una serie de frecuencias se trabajará los datos de la **Tabla N°17**:

intervalo de clase	f_i	x_i	$(x_i - \bar{x})^2 * f_i$
3 - 9	62	6	3968
9 - 15	40	12	160
15 - 21	31	18	496
21 - 27	15	24	1500

27 - 33	12	30	3072
33 - 39	5	36	2420
Total	165		11616

$\bar{x} = \frac{2310}{165} = 14$ $s = \sqrt{\frac{11616}{164}} = 8.42$

COEFICIENTE DE VARIACIÓN

Las tres medidas de variabilidad enunciadas precedentemente son medidas de variabilidad absoluta. El coeficiente de variación es una medida de variabilidad relativa. Expresa la desviación estándar como un porcentaje de la media.

$CV\% = \frac{S}{\bar{x}} * 100$

En el ejemplo de la serie simple:

$CV\% = \frac{4.36}{27} * 100 = 16.15\%$

En el ejemplo de la serie de frecuencias:

$CV\% = \frac{8.42}{14} * 100 = 60.14\%$

Uso de la calculadora científica para el cálculo de Medidas de Posición y Dispersión.

Seguir las siguientes instrucciones:

- Debe procurar que la calculadora se encuentre en disposición para efectuar cálculos estadísticos. Para ello en la parte superior de la pantalla debe aparecer la notación SD. En algunas calculadoras esto se consigue haciendo MODE.
- Debe cerciorarse de que no hay nada acumulado. Para ello debe pulsar la tecla n. En algunas calculadoras esto se consigue haciendo INV 6. Si sale 0 en la pantalla se está en condiciones de acumular los datos. Si no hay que borrar lo que hay en memoria haciendo INV AC.
- Acumulación de datos:
  1er dato y se aprieta M+
  2do dato y se aprieta M+
  Así sucesivamente hasta haber cargado todos los datos.
- Pulsando INV 6 obtenemos el número de datos introducidos; INV 7 la media aritmética.
- Si se tiene una serie de frecuencias la acumulación de datos se debe hacer así:
  1er dato x primera frecuencia M+
  2° dato x segunda frecuencia M+
  y luego se procede como en la serie simple para obtener la media aritmética.
- Para obtener la desviación estándar se aprieta INV9(si se trabaja con muestras) o INV8 (si se trabaja con población)

CAPITULO III

PROBABILIDADES Y DISTRIBUCIONES DE PROBABILIDADES

Probabilidad

Es la posibilidad de ocurrencia de un hecho. Matemáticamente se mide con un número que va desde 0 hasta 1 o, si así se lo desea, en por ciento desde 0 a 100%.

Probabilidad y Estadística

Como ya se vió, en la Estadística Descriptiva, se hace referencia a los datos que se tienen en la mano. Cuando se quiere ir más allá de los datos disponibles, es necesario inferir o sea utilizar la Estadística Inferencial. Como ella infiere el todo (población) a partir de la información que da una parte de ese todo (muestra), el conocimiento que adquiere es incompleto y por lo tanto no “totalmente cierto” es decir, se debe trabajar con probabilidades. Por ello, antes de estudiar las aplicaciones de la Estadística Inferencial es necesario estudiar probabilidades.

Experimentos aleatorios. Espacio muestral. Eventos.

Las probabilidades se aplican a los experimentos aleatorios que son aquéllos que, repetidos bajo idénticas condiciones, no arrojan un único resultado sino un conjunto de ellos. Ese conjunto de los resultados posibles de un experimento aleatorio recibe el nombre de espacio muestral (M) y cada uno de los resultados es un evento simple.

Un ejemplo muy sencillo es el del experimento aleatorio consistente en arrojar un dado. El espacio muestral M es: M= $\{1,2,3,4,5,6\}$ o sea el conjunto de resultados posibles al arrojarlo. Cada uno de esos resultados es un evento simple. Un evento compuesto es el resultado de la unión de uno o más simples. Por ejemplo, el evento: obtener un nº par es
P = $\{2,4,6\}$.

El diámetro de un árbol elegido al azar entre todos los árboles de una plantación también constituye un experimento aleatorio. En este caso. el espacio muestral no es finito y lo podemos representar como M = $\{x / x \in \Re \wedge 10 \leq x \leq 60\}$
lo que quiere decir que el diámetro del árbol elegido puede tomar cualquier valor entre 10 y 60 cm (ambos incluídos).

Definición clásica de probabilidad (probabilidad a priori)

Es el cociente entre el número de casos favorables y el número de casos posibles (todos son igualmente posibles).

Ejemplo: si se arroja un dado perfecto, cada una de las caras tiene igual probabilidad de ocurrencia, o sea que P = 1/6.

Definición de probabilidad frecuencial (probabilidad a posteriori)

Las probabilidades se aproximan después de realizar la experiencia. Por ejemplo, para saber cuál es la probabilidad de obtener el as con un dado determinado, se arroja el dado 600 veces en las cuales se obtienen 113 veces un as.

La probabilidad de obtener un as con ese dado es estimada por la frecuencia relativa = 113/600 = 0.1883. Por lo que, en símbolos se puede escribir:

$$P(As) \approx \frac{f_{(As)}}{\sum f} = f_{r(As)}$$

Teorema de la suma de probabilidades

Sean A y B dos eventos del espacio muestral M generado por un experimento aleatorio. El teorema de la suma de probabilidades dice que la probabilidad de la unión de A y B es la suma de las probabilidades menos la probabilidad de la intersección. En símbolos:

Sean A y B $\subset M$ entonces $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Teorema del producto de probabilidades

La probabilidad de la intersección es el producto de las probabilidades:

P(A∩B) = P(A)*P(B)   en caso de independencia
= P(A)*P(B/A)   en caso de eventos dependientes

Variable aleatoria

Es aquélla cuyos valores están determinados por los resultados de un experimento aleatorio.

Distribuciones de probabilidades de variable aleatoria discontinua

La siguiente es la distribución de la variable aleatoria X_i = nº de puntos obtenidos al arrojar un dado perfecto o sea que todas sus caras son igualmente posibles:

x_i	1	2	3	4	5	6	Total
$P(x_i)$	1/6	1/6	1/6	1/6	1/6	1/6	1

Obsérvese que se cumplen dos condiciones que son necesarias para que un conjunto de pares ordenados (x,y) sea considerada una distribución de probabilidades:

1) para cada valor de **x** le corresponde un único valor de **y** que es un valor de probabilidad (no negativo y menor o igual a 1),
2) $\sum P(x_i) = 1$.

Esta distribución recibe el nombre de uniforme, es una distribución de variable aleatoria discontinua y sus parámetros son los valores mínimo (a) y máximo (b) que puede tomar x. Esto se indica como $X \sim U(a, b)$.

Otra distribución de variable aleatoria discontinua, muy utilizada es la distribución Binomial.

La variable x toma los valores 0, 1, 2, 3, ... , n. (donde n es finito y bien determinado). Se puede considerar que la distribución binomial es la repetición de n pruebas independientes (por ejemplo poner a germinar 4 semillas). La función de probabilidades es:

$$P(x) = C_n^x p^x q^{n-x}$$

donde C_n^x son las combinaciones de n elementos tomadas de a x,

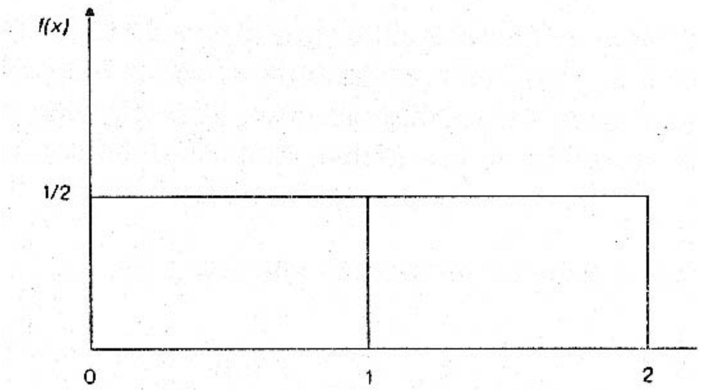
p= probabilidad de éxito en una sola prueba,
q = 1- p = probabilidad de fracaso.

La combinaciones se calculan como sigue: C_n^x = (n.(n-1). (n-2)...(n-x+1))/x!
Los parámetros que definen a la distribución Binomial son n y p

Distribuciones de probabilidades de variable aleatoria continua

En estas distribuciones no es posible calcular la probabilidad en puntos sino que hay que hacerlo en intervalos. Recuérdese que en las variables discontinuas las probabilidades de intervalos se obtenían sumando las probabilidades que corresponden a cada punto o valor de la variable. En variables continuas, los valores que puede tomar la variable son infinitos por lo que es necesario hacer una suma infinita es decir una integral. En las variables continuas, la probabilidad de un intervalo se obtiene integrando la función de densidad.

Ejemplo: la distribución rectangular X ~ R (0,2).
Esta es una distribución rectangular (todos sus puntos tienen igual densidad de probabilidad) que se extiende desde 0 a 2. El gráfico de su función de densidad es el siguiente:



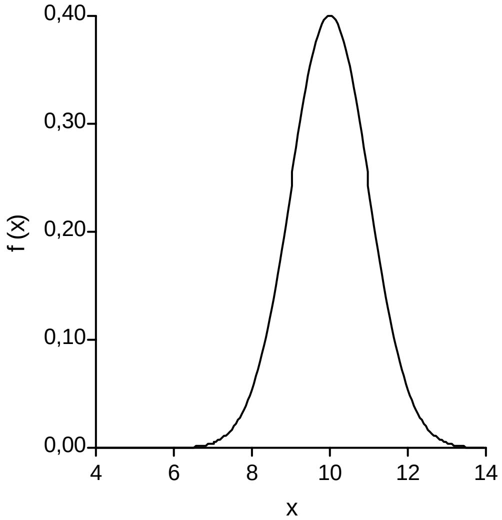
en el que se puede observar que la función de densidad
f(x) = 1/2
La probabilidad de encontrar valores de variables entre 1 y 2 se encuentra integrando la función de densidad entre esos límites.
La integral entre esos límites corresponde al área bajo de la curva entre los mismos.

La distribución Normal

Si una variable es continua, varía desde $-\infty$ hasta $+\infty$ y su función de densidad es:

$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, se dice que x tiene distribución normal con parámetros μ y σ (media aritmética y desviación estándar). Esto se simboliza como sigue : X ~ N (μ , σ)

Su gráfica es la siguiente:



La distribución normal presenta las siguientes características:

1) Presenta un máximo en x = μ , por lo tanto Mo = μ
2) Es simétrica y su eje de simetría es $f(\mu)$, por lo que se deduce que Md = μ = Mo.
3) Tiene dos puntos de inflexión ubicados en $x = \mu \pm \sigma$
4) Toda transformación lineal de x da otra distribución normal.
5) Algunos sectores usados de la función son:

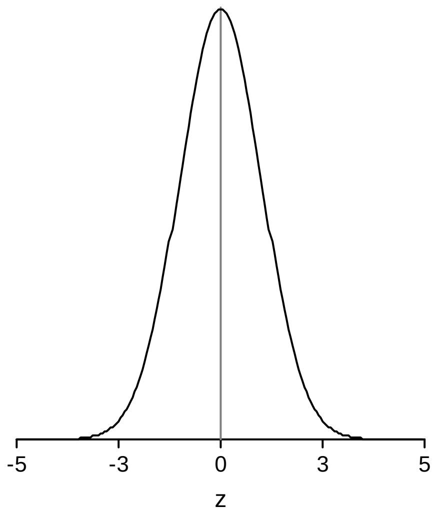
$x = \mu \pm \sigma$ corresponde aproximadamente al 68 % central
$x = \mu \pm 2\sigma$ corresponde aproximadamente al 95 % central
$x = \mu \pm 3\sigma$ corresponde aproximadamente al 99 % central

6) f(x) se acerca asintóticamente al eje x o sea que f(x) > 0.
7) Por ser función de densidad, el área bajo de la curva es $\int_{-\infty}^{+\infty} f(x)dx = 1$

Para calcular la probabilidad de un intervalo en la distribución normal, por tratarse de una variable continua, debe hacerse mediante la integración de la función de densidad, lo cual equivale a calcular el área bajo de la curva. Considérese por ejemplo que el peso específico de la madera de una especie tiene distribución normal con media μ = 0.6 kg/dm^3 y desviación estándar σ = 0.1 dm^3. La probabilidad de obtener muestras de esa madera con valores de densidad comprendidos entre 0.75 y 0.5 (P(0.5<x<0.75)) se obtiene integrando la función de densidad f(x), (en la cual se debe reemplazar los valores correspondientes de μ y σ por 0.6 y 0.1 respectivamente)entre los límites 0.5 y 0.75.

La distribución normal estándar

Usando la propiedad que dice que la transformación lineal $z = \dfrac{x - \mu}{\sigma}$ conduce a una distribución también normal, cuyos parámetros son $\mu_z = 0$ y $\sigma_z = 1$, se obtiene una nueva distribución que se conoce con el nombre de distribución normal estándar o normal 0,1 y se la describe como
Z ~ N(0,1) cuya representación gráfica es la siguiente:



Tablas de la distribución normal

El cálculo de probabilidades en la normal involucra el cálculo de integrales que son muy engorrosas de resolver manualmente. Por ello, las integrales están tabuladas para una distribución normal que es la estándar.

Se disponen de dos tablas de la distribución normal: la tabla 1 o tabla de “1 cola” y la tabla 2 o “tabla de 2 colas”.

Tabla de “1 cola”

En ella, los valores de probabilidad se encuentran en el cuerpo de la tabla y los valores de z se forman utilizando la primera columna y la primera fila (es decir en lo que se conoce como matriz de la tabla). En esta tabla es importante considerar el signo de z.

Como su nombre lo indica, para el valor de z considerado, da el valor del área bajo de la curva desde menos infinito hasta z. Por ejemplo si z = -2.1 la tabla da $P(z < -2.1) = 0.0179$.

Ejemplo : Una población de pesos de semillas en gr tiene distribución normal con media y desviación estándar (μ y σ) de 2 y 0.2 gr. respectivamente. ¿Cuál es la probabilidad de semillas que pesen menos de 2.3 gr.?

En símbolos, la probabilidad buscada es $P(x \leq 2.3)$

Para solucionar esto es necesario pasar de la normal que nos interesa a la normal estándar. Esto se consigue mediante el siguiente cambio de variable:

$z = \dfrac{x - \mu}{\sigma}$ en este caso x = 2.3 , $\mu = 2$ y $\sigma = 0.2$ por lo que
$z = (2.3 - 2)/0.2 = 1.5$.

$P(x \leq 2.3) = P(z \leq 1.5) = 0.9332$

Tabla de "2 colas"

Esta tabla está construída de manera diferente: en el cuerpo de la tabla se encuentran los valores de z y en la matriz, los valores de probabilidad. El nombre de "dos colas" se refiere a que la tabla da el área de las dos colas simétricas. Por ejemplo para z = 0.51 corresponde P = 0.61. Esta probabilidad es la suma de $P(z < -0.51)$ y $P(z>0.51)$.

Esta tabla es útil cuando se quieren definir intervalos centrales simétricos que corresponde a un porcentaje determinado de la población.

Ejemplo: En la población de pesos X ~ N(2 ; 0.2) ¿cuál es el intervalo que corresponde al 95 % central de la población?

En la distribución de z, el 95 % central de la población corresponde al intervalo que va desde -1.96 a +1.96, o sea $\pm$1.96. ¿Cómo se busca este valor en la tabla de dos colas? Es muy simple, si se desea que en el centro esté el 95 % o, en tanto por uno, 0.95, entonces en las colas debe quedar el 0.05. Buscando para P= 0.05 se encuentra z = 1.959964 que se aproxima a 1.96.

Ya se determinó el intervalo en z, ¿cómo se pasa a la normal con media 2 y desviación estándar 0.2? Se debe hacer el cambio inverso de variable: $z = \frac{x - \mu}{\sigma}$ entonces $x = \mu \pm z\sigma$.

Para indicar que el intervalo corresponde a un porcentaje central determinado se acostumbra a llamar α a lo que queda en las colas, o sea a la probabilidad con la que se entra en la tabla. De este modo, el intervalo que corresponde a un porcentaje central de la población de $(1 - \alpha)$ % es:

$$x = \mu \pm z_\alpha \sigma$$

Volviendo entonces al ejemplo, por ser z_α= 1.96, el intervalo que corresponde al 95 % de la población de pesos de frutos es:

$$x = 2 \pm 1.96\ 0.2 = 2 \pm 0.392 .$$

El intervalo entonces va desde 1.608 gr. a 2.392 gr.

CAPITULO IV

TEORÍA ELEMENTAL DE MUESTREO. ESTIMACIÓN DE LA MEDIA POBLACIONAL POR INTERVALO. ESTIMACIÓN DE LA PROPORCIÓN POBLACIONAL.

Generalmente sucede que es imposible estudiar mediante censo el objeto de nuestro estudio: por ejemplo medir las alturas de todas las plantas de un vivero comercial o examinarlas a todas para ver su calidad, o medir todos los árboles de una plantación. Es por ello necesario estudiar una parte de él o sea estudiar una muestra. Los dos conceptos siguientes son importantes para este tema.

Población: Es el todo del cual será extraída una muestra. Está constituida por N unidades muestrales. Las unidades muestrales pueden ser, de acuerdo a lo que se esté estudiando: parcelas, árboles, hojas, insectos, etc.

Muestra: Subconjunto de la población constituida por n unidades muestrales.

Para que una muestra pueda ser estudiada estadísticamente debe ser seleccionada al azar: debe ser **aleatoria**.

Aquí es conveniente efectuar la siguiente observación: una unidad muestral puede estar definida por la naturaleza (como es el caso de un árbol, una hoja, una rama, una plántula) o debe ser definida por el técnico: parcela (tamaño y forma). A veces es conveniente o más cómodo elegir un grupo de unidades muestrales, en este caso se habla de muestreo por conglomerados.

La teoría elemental del muestreo permite conocer que sucede cuando se extraen muestras de una población y por ende, conocer las distribuciones muestrales (distribuciones que se originan en el muestreo) y si las muestras han sido elegidas al azar, son distribuciones de variables aleatorias. Su conocimiento permite entonces, a partir de una muestra, inferir a la población.

Parámetros y estimadores

Los parámetros son valores constantes bajo determinadas condiciones. En una población, por ejemplo, la media aritmética y la desviación estándar constituyen parámetros de tendencia central y de variabilidad respectivamente. En las poblaciones, sus parámetros sólo pueden ser conocidos cuando se efectúan censos:

N = tamaño de la población

La media aritmética de la población es $\mu = \frac{\sum_{i=1}^{i=N} x_i}{N}$ y su desviación estándar es $\sigma = \sqrt{\frac{\sum_{i=1}^{i=N} (x_i - \mu)^2}{N}}$.

Pero como casi siempre es imposible realizar un censo, se debe trabajar con los datos de una muestra. El tamaño de la muestra se designa con **n**. Los valores que se calculan en la muestra, como su media y desviación estándar, se llaman **estimadores ó estadísticos** y se calculan con las siguientes fórmulas:

$\overline{x} = \frac{\sum_{i=1}^{i=n} x_i}{n}$; $S = \sqrt{\frac{\sum_{i=1}^{i=n} (x_i - \overline{x})^2}{n - 1}}$, ambos son muy buenos estimadores de los parámetros a los cuales estiman : μ y σ.

Muestreo al azar simple

En este tipo de muestreo todas las N unidades muestrales de la población tienen igual probabilidad de integrar la muestra. Consideraremos únicamente el muestreo en población infinita el cual es equivalente al muestreo
con reposición en población finita. Para obtener una muestra al azar simple es necesario utilizar un procedimiento que garantice la selección aleatoria y con igual probabilidad.

El procedimiento cambia según el problema que se estudie. Cuando es posible identificar a todas las unidades muestrales mediante un número, se seleccionan luego números al azar de tal manera que todos tengan igual probabilidad de ser seleccionados (distribución uniforme). Esto se puede hacer utilizando una tabla de números aleatorios o simplemente usando la función random (azar en inglés) de las calculadoras (RND).

Supóngase que se tiene una población de tamaño
N = 1000 y se desea extraer una muestra de tamaño n = 10. La función random generalmente trabaja de la siguiente manera: Se da un número inicial que hace de semilla y la calculadora internamente, con esa semilla, calcula un número al azar en una distribución uniforme que va desde 0 hasta 1. El número aleatorio obtenido debe ser luego llevado hasta N mediante un simple producto.

En el ejemplo de la población de N = 1000, supóngase que el primer número elegido es 0.752. Multiplicado por N da 752 lo que significa que la unidad muestral identificada con el número 752 debe ser incluida en la muestra. De esta manera se repite el procedimiento hasta completar la muestra.

Una vez completada la muestra, se calculan los estimadores $\bar{x}$ y S

$\bar{x} \Rightarrow \mu$ y $S \Rightarrow \sigma$ en donde la flecha quiere decir "estima". En el muestreo al azar simple ambos son estimadores insesgados.

Insesgado significa que en promedio son iguales al parámetro que estiman. Por ejemplo si sacamos todas las nuestras posibles de tamaño **n** de una población, calculamos en cada una de ellas $\bar{x}$, la media aritmética de las $\bar{x}$, será igual a μ.

Estimación por punto

Las estimaciones puntuales consisten en estimar el parámetro poblacional mediante un único valor, el valor del estimador muestral.

Ejemplo:

Sea la siguiente muestra de la altura de las plántulas de una determinada especie que produce un vivero :

x_i (altura en cm) : 15, 25, 7, 16, 18, 19, 20, 21.

Estimar por punto a μ.

$\bar{x}$ = 17.625 cm por lo tanto la estimación puntual de μ es :

$$\hat{\mu} = 17.625 \text{ cm}$$

Las estimaciones puntuales son de valor relativo ya que, en la práctica, extraemos una sola muestra y su $\bar{x}$ puede estar muy alejada de μ , lo cual no lo podemos saber ya que no conocemos a la población. Lo único que sabemos es que $\bar{x}$ es un estimador insesgado de μ , es decir, que en promedio $\bar{x}$ es igual a μ.

Distribución de medias muestrales

Si se eligen al azar muestras de tamaño **n** y en cada una de ellas se calcula la media aritmética $\bar{x}$, al ser las muestras elegidas al azar, $\bar{x}$ es también una **variable aleatoria**.

¿Qué distribución tiene la variable aleatoria $\bar{x}$?

a) Si la población original es normal : $x \sim N(\mu, \sigma)$ entonces

$\bar{x} \sim N(\mu, \sigma_{\bar{x}})$

$\sigma_{\bar{x}}$ es la desviación estándar de la población de $\bar{x}$ y se calcula como sigue:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Esta última fórmula indica que la variabilidad de la población de medias muestrales depende directamente de la variabilidad de la población original e inversamente de la raíz cuadrada del tamaño de muestra **n**.

b) Si la población original no es normal la solución es trabajar con muestras mayores de 30 que, a los fines de la práctica, aseguran una aproximación suficiente a la distribución normal.

Como en cualquier variable normal, si calculamos

$z = \dfrac{\bar{x} - \mu}{\sigma_{\bar{x}}}$, la variable z tiene distribución normal estándar.

Conocidos la forma y los parámetros de la distribución de $\bar{x}$, es posible hacer estimaciones por intervalo y conocer la confianza del intervalo .

Variables contínuas: Estimación de μ por intervalo siendo σ conocido

En base al conocimiento de la distribución de $\bar{x}$ se puede estimar a μ por intervalo. El intervalo se genera de la siguiente manera:

$$\hat{\mu} = \bar{x} \pm z_{\alpha}\, \sigma_{\bar{x}}$$

El producto de z_{α} por $\sigma_{\bar{x}}$ se denomina error de estimación y se designa por E. Al sumar y restar E a $\bar{x}$ se obtiene un valor superior y otro inferior que reciben el nombre de límites fiduciales y constituyen los límites del intervalo de confianza. Si $\alpha = 0.05$ (o sea el 5 %) la confianza es su complemento a 100 o sea 100 - 5 = 95 % de confianza. Esto significa que la probabilidad de que el intervalo así construido contenga a μ es del 95 %.

Lo deseable es que los intervalos sean lo más pequeños posibles (E sea chico). Esto se puede conseguir de dos maneras:

a) Disminuyendo z, lo que significa que α aumenta y esto no conviene porque disminuye la confianza, ó,

b) Disminuyendo $\sigma_{\bar{x}}$. En el caso de que las unidades muestrales estén ya predefinidas por la naturaleza, la única forma de lograrlo es aumentando el tamaño de muestra **n** porque σ es un valor que no se puede modificar ya que constituye una característica propia de la población que se estudia. En el caso de parcelas, aumentando su tamaño se puede disminuir el coeficiente de variabilidad o sea la variabilidad de la población. De esta manera es posible tener influencia sobre σ y elegir el tamaño de parcela más conveniente.

Error de estimación relativo o porcentual (E %)

Muchas veces es conveniente expresar el error de estimación E en relación a la media $\bar{x}$, en ese caso el error se denomina porcentual o relativo y se lo calcula como:

$$E\% = \frac{E}{\bar{x}}\,100 = \frac{z_\alpha\,\sigma_{\bar{x}}}{\bar{x}}\,100 = \frac{z_\alpha\,\sigma}{\bar{x}\sqrt{n}}\,100 = \frac{z_\alpha\,CV\%}{\sqrt{n}}$$

Cálculo del tamaño de la muestra n para cometer un error E determinado

El examen de la fórmula anterior permite observar que es factible calcular el tamaño de la muestra **n** para obtener un determinado error de estimación E (ya sea absoluto o porcentual). Con unos simples pasos algebraicos se puede demostrar que:

$$n = \left(\frac{z_\alpha\,\sigma}{E}\right)^2 = \left(\frac{z_\alpha\,CV\%}{E\%}\right)^2$$

Variables contínuas: Estimación de μ por intervalo siendo σ desconocido. La distribución t de Student.

La situación que hemos descripto es la menos corriente ya que al ser σ un parámetro poblacional no es conocido. Es entonces necesario estimarlo con el valor muestral S, por lo que $\sigma_{\bar{x}}$ también será estimado por $S_{\bar{x}} = \frac{S}{\sqrt{n}}$. En esta situación, y bajo la condición de que la variable x cuya media poblacional se está estimando tenga distribución normal, la variable $\frac{\bar{x} - \mu}{S_{\bar{x}}}$ se denomina t y tiene una distribución que no es la Z (normal estándar) sino que se llama $t_{(n-1)}$.

Ella constituye una familia de distribuciones con forma simétrica y acampanada muy similar a la normal. Su semejanza aumenta a medida que aumentan los grados de libertad (n - 1). Los grados de libertad son el parámetro que la define. Se dispone de una tabla de t, es una tabla de una cola.

Utilizando t, la forma de calcular el intervalo del (1 - α) % de confianza de la media poblacional μ es:

$$\mu = \bar{x} \pm t_{(n-1)\alpha}\,S_{\bar{x}}$$

En el ejemplo de la estimación de la media poblacional de las alturas de las plántulas de un vivero la muestra, de tamaño n = 8 que tiene $\bar{x}$ = 17.625 cm y S =, 5.2898, el intervalo del 95 % de confianza para la estimación de μ se calcula como sigue:

13.2019 cm

$$\hat{\mu} = 17.625 \pm 2.365 \cdot \frac{5.2898}{\sqrt{8}} = 17.625 \pm 4.4231 = $$

22.0481 cm

el valor de t se busca en la tabla para 7 grados de libertad y α = 0.025 ya que se trata de una tabla de una cola por lo que para tener 5 % en las dos colas, debe ser cada una de 2.5 %.

De igual forma, el cálculo del tamaño de la muestra necesaria para cometer un error determinado se hace utilizando fórmulas similares a las presentadas para el caso de tener σ conocido, pero cambiando σ por su estimador S y z_α por $t_{(n-1)\alpha}$

$$n = \left(\frac{\mathrm{t\,S}}{\mathrm{E}}\right)^2 = \left(\frac{\mathrm{t\,CV\%}}{\mathrm{E\%}}\right)^2$$

Variables cualitativas: Estimación de P por intervalo (caso en que la unidad de muestreo es el individuo)

Muchas veces interesa estimar a la proporción poblacional P. Sea el caso de la proporción de plantas enfermas en un vivero, P . La misma es el cociente entre el número de plantas enfermas (A) sobre el total de plantas del vivero (N)

$$P = \frac{\mathrm{A}}{\mathrm{N}}$$

La estimación puntual se hace mediante una muestra de tamaño **n** con el cociente del número de plantas enfermas de la muestra (a) y el tamaño de la muestra : $p = \frac{\mathrm{a}}{\mathrm{n}}$.

Cuando el tamaño de la muestra es lo suficientemente grande (ver Tabla 1) se puede usar una aproximación a la distribución normal para hacer estimaciones por intervalo.

Tabla 19: Valores mínimos de n (según p) para uso de la aproximación normal (Cochran 1974)

p	n
0.5	30
0.4	50
0.3	80
0.2	200
0.1	600
0.05	1400

$$\hat{P} = p \pm \left(z_\alpha\ S_p + \frac{1}{2n} \right) \qquad \text{donde} \qquad S_p = \sqrt{\frac{\mathrm{p\,(1\text{-}p)}}{\mathrm{n}}}$$

1/2n se conoce como corrección por continuidad.

Ejemplo: En una muestra de 200 tablas de un aserradero se encuentran 42 con fallas. Estimar por intervalo del 95 % de confianza a la proporción de defectuosas de la población.

$p = 0.21,$ $S_p = \sqrt{\frac{0.21 \cdot 0.79}{100}} = 0.0407$

$\hat{P}$ = 0.21 ± (1.96 . 0.0407 +1/400) = 0.21 ± 0.0823 .

Sumando y restando el error de estimación obtenemos los límites fiduciales superior e inferior: 0.2923 y 0.1277. Se puede decir entonces que la proporción de tablas defectuosas está entre 0.29 y 0.13 con una confianza del 95 %.

Variables cualitativas: Estimación de P por intervalo (caso de muestreo por conglomerados)

En la mayoría de las situaciones, el muestreo al azar simple es incómodo y poco práctico para su aplicación. Por ejemplo en viveros, es mucho más práctico seleccionar conjuntos de plantas para estudiar, por ejemplo el porcentaje de atacadas. O, en la determinación del poder germinativo, tomar conjuntos de semillas.

Este tipo de muestreo se denomina por conglomerados y se distinguen en él dos casos:

a) Conglomerados de igual tamaño y,

b) Conglomerados de distinto tamaño

a) Conglomerados de igual tamaño: en este caso, es posible calcular en cada conglomerado un valor p_i , donde i simboliza al iésimo conglomerado y varía desde 1 a n.

En este caso el intervalo para estimar a P con una confianza de 1- α es:

$$\hat{P} = \bar{p} \pm t_{(n-1)\alpha} S_{\bar{p}}$$

en donde $\bar{p}$ y $S_{\bar{p}}$ son la media aritmética y la desviación estandar de la media aritmética de las proporciones y se calculan de la manera habitual:

$$\bar{p} = \frac{\sum p_i}{n} \quad ; \quad S_{\bar{p}} = \frac{S_p}{\sqrt{n}} \quad ; \quad S_p = \sqrt{\frac{\sum (p_i - \bar{p})^2}{n-1}}$$

Ejemplo: Supóngase que se ponen a germinar 10 grupos, cada uno de ellos de 50 semillas para estudiar el poder germinativo de las semillas de una especie. Los resultados obtenidos figuran a continuación, en donde x_i es el número de semillas que germinaron y p_i la proporción (o poder germinativo en tanto por uno) obtenido con el cociente: x_i / 50.

Gru.	1	2	3	4	5	6	7	8	9	10
x_i	42	36	48	45	39	42	32	35	42	43
p_i	0.84	0.72	0.96	0.9	0.78	0.84	0.64	0.7	0.84	0.86

$$\overline{p} = \frac{\sum p_i}{n} = 0.808; \; S_p = \sqrt{\frac{\sum (p_i - \overline{p})^2}{n-1}} = 0.0976;$$

$$S_{\overline{p}} = \frac{S_p}{\sqrt{10}} = \frac{0.0976}{3.1623} = 0.0309$$

El valor de $t_{(9)0.05}$ es 2.262. En consecuencia, la estimación por intervalo de P es:

$$\hat{P} = 0.808 \pm 2.262 * 0.0309 = 0.808 \pm 0.0698$$

Redondeando, entre 0.74 y 0.88

b) Conglomerados de distinto tamaño: Es un caso común en la práctica del inventario forestal desde que tanto las parcelas como las estaciones de muestreo angular comprenden un número variable de árboles. En este caso, para estimar P se usa un estimador de razones, en el que se tiene en cuenta el hecho que los datos provengan de conglomerados de distinto tamaño. Si se designa con y_i al número de individuos que presentan el atributo buscado en el conglomerado iésimo , con x_i al tamaño del conglomerado, $\hat{P}$ y S_P se calculan como sigue:

$$\hat{P} = \frac{\sum y_i}{\sum x_i} = \frac{\overline{y}}{\overline{x}} \quad ; \quad S_P = \sqrt{\frac{1}{\overline{x}}\left(\frac{S_y^2 + \hat{p}^2 S_x^2 - 2PS_{yx}}{n}\right)}$$

donde: $\overline{X}$: número promedio de árboles por parcela;
n : número de parcelas;
$\hat{P}$: es el estimador puntual del valor poblacional P;
S^2_y , S^2_x son las variancias de “y” y “x “ respectivamente,
S_{yx} es la covariancia entre x e y.

Las fórmulas de las variancias ya son conocidas por el lector. La covariancia se calcula como sigue

$$S_{yx} = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{n-1}$$

Se presentan como ejemplo a 10 parcelas de 300 m^2 con la información del número total de árboles por parcela (x_i) y el número de árboles dañados por parcela (y_i). Estimar por intervalo del 95% de confianza a la proporción P de árboles dañados.

Tabla 19. Valores del número total de árboles por parcela (x_i), número de árboles dañados por parcela (y_i) y columnas auxiliares para el calculo .

Parcela i	x_i	y_i	X_i	Y_i	X_i*Y_i	X^2_i	Y^2_i

1	31	6	6	0.8	4.8	36	0.64
2	28	6	3	0.8	2.4	9	0.64
3	20	5	-5	-0.2	1.0	25	0.04
4	23	4	-2	-1.2	2.4	4	1.44
5	29	5	4	-0.2	-0.8	16	0.04
6	14	3	-11	-2.2	24.2	121	4.84
7	18	5	-7	-0.2	1.4	49	0.04
8	25	6	0	0.8	0	0	0.64
9	32	6	7	0.8	5.6	49	0.64
10	30	6	5	0.8	4	25	0.64
Total	250	52	0	0	45	334	9.6

$$\bar{x} = 25 \quad : \quad \bar{y} = 5.2 \quad ; \hat{p} = \frac{\bar{y}}{\bar{x}} = \frac{5.2}{25} = 0.208$$

$$S_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{334}{9} = 37.1111$$

$$S_y^2 = \frac{\sum (y_i - \bar{y})^2}{n-1} = \frac{9.6}{9} = 1.0667$$

$$S_{yx} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{45}{9} = 5$$

$$S_P = \sqrt{\frac{1}{\bar{x}^2}\left(\frac{S_y^2 + \hat{p}^2 S_x^2 - 2PS_{yx}}{n}\right)} =$$

$$= \sqrt{\frac{1.0067 + 0.208^2 * 37.1111 - 2 * 0.208 * 5}{25^2 * 10}} = 0.0097$$

El intervalo se construye con $t_{(n-1)0.05}$ de la siguiente manera:

$$\hat{P} = 0.208 \pm 2.262 * 0.0097 = 0.208 \pm 0.0220$$

Los límites superior e inferior son respectivamente 0.23 y 0.186. Por lo tanto, el porcentaje de plantas dañadas se encuentra entre 18.6% y 23% con una confianza del 95%.

Muestreo al azar estratificado:

En todos los ejemplos presentados la selección de la muestra (mediante individuos o conglomerados) se hizo sin restricciones o sea al azar simple. Si en la población es posible distinguir partes (subpoblaciones) con características definidas e identidad propia y que se diferencian de las otras partes, es muy conveniente utilizar el muestreo estratificado. Cada una de las subpoblaciones recibe el nombre de estrato y el muestreo se realiza al azar simple en cada uno de los estratos.

El muestreo estratificado es muy eficiente, es decir que sus errores de estimación son menores que los del azar simple siempre y cuando se haya estratificado correctamente (esto es así cuando la variabilidad dentro de los estratos es menor que la de la población sin estratificar).

Este muestreo escapa a los contenidos mínimos de esta asignatura pero es importante que el técnico forestal conozca su existencia y podrá utilizarlo si fuera necesario, con la ayuda de un texto básico de Estadística o de muestreo.

CAPITULO V

GUIA DE EJERCITACION

1.- ¿Qué se entiende por variable y de que tipos pueden ser las variables que nos interesan?

2.- De 5 ejemplos de cada tipo de variable.

3.- Escriba tres series simples de datos (una para cada tipo de variable), n = 10.

4.- ¿Qué se entiende por serie de frecuencias?

5.- Escriba tres series de frecuencias (una para cada tipo de variable) de tal modo que en cada una de ellas $\sum f_i = 60$. Para una mayor prolijidad, presente las series de frecuencias en tablas.

6.- ¿Qué tipos de tablas son las que presentó en el ejercicio anterior?

7.- ¿Qué se entiende por tabla de doble entrada? Dé un ejemplo de tabla de doble entrada.

8.- En el Anexo 1 de esta guía se encuentra una planilla de campo de la medición de una parcela de muestreo en una plantación de álamos en Santiago del Estero. En la columna titulada observaciones se encuentra el estado sanitario (S = sano, E = enfermo). Agrupe en una serie de frecuencias según la variable estado sanitario a los árboles estudiados y presente los resultados del agrupamiento en una tabla.

9.- Represente mediante el gráfico apropiado a la tabla obtenida en el ejercicio anterior.

10.- En la planilla de campo del Anexo 1 la columna titulada nb contiene la información del número de brotes (ramitas) en el fuste. Agrupe a los 40 árboles según esta variable y presente la tabla correspondiente.

11.- Represente mediante el gráfico apropiado a la tabla obtenida en el ejercicio anterior.

12.- Agrupe en serie de frecuencias a los árboles del Anexo 1 según la variable dap.

13.- Represente mediante el gráfico apropiado a la tabla obtenida en el ejercicio anterior.

14.- ¿Cómo es la relación entre el diámetro y la altura total en los árboles del Anexo 1?

15.- Mediante un gráfico de sectores represente las superficies forestadas con distintas especies en Entre Ríos al año 1995 según los datos de la tabla que sigue:

Superficie forestada (has) en Entre Ríos al año 1995

CONÍFERAS	EUCALYTUS	SALICÁCEAS	TOTAL
9.197	54.470	20.295	83.962

Fuente: SAGPYA, 1998

16.- Utilice superficies para representar las superficies logradas mediante el régimen de promoción de plantaciones forestales en 1995 en corrientes (13.440 has) y en Entre Ríos (885 has) según SAGPYA, 1998

17.- Presente fotocopias de dos tablas y dos gráficos tomados de cualquier publicación. En cada uno de ellos haga los comentarios que estime convenientes: interpretación, errores, carencias, redundancias, etc.
Nota: en todas las tablas y gráficos no olvide poner el título completo, fuente, nominar ejes, referencias y todo lo que sea necesario para una completa comprensión del que lee.

18.- Durante la primera semana del mes de Abril de 1999 se registraron las siguientes temperaturas diarias, medidas en grados centígrados:
16 ; 16.1 ; 16.8 ; 18.8 ; 18.9 ; 18.3 ; 17.8
¿ Cuál es la temperatura promedio de esa semana?

19.- Para llenar bolsitas con tierra en un vivero , se dispone de 10 operarios . El rendimiento de los mismos en un día de trabajo es el siguiente:

Nº de bolsas por operario (x_i)	Nº de operarios (f_i)
250	1
125	2
250	3
300	4
Total operarios	10

¿ Cuál es el promedio de bolsas que se llenan por día, por operario?
¿ Cuál es el total de bolsas que se llenaran en una semana , por operario? Considere que son cinco días hábiles de trabajo.
Pruebe la propiedad que nos dice que la suma de los desvíos con respecto a la media es cero.

20.- En una plantación se considera la siguiente variable :
x_i : número de árboles atacados por insectos en una parcela
y se cuentan las parcelas con x_i árboles atacados obteniéndose así las frecuencias f_i

Obtenemos la siguiente serie agrupada

x_i	f_i
0	80
1	60
2	30
3	25
4	10

Calcule el promedio de árboles atacados .

21.- La siguiente es una tabla donde se muestra las alturas de álamos en una parcela de una plantación del Dto. Capital, Sgo. del Estero.

clases de altura (m) x_i	f_i
1 - 4	1
4 - 8	3
8 - 12	4
12 - 16	14
16 - 20	5
Total árboles	27

Calcule la altura promedio de la parcela.

22.- Todos los ejercicios que resolvió anteriormente en forma manual deberá resolverlos ahora usando calculadora científica.

23.- Calcule la Mediana en las siguientes series simples:
5 ; 13 ; 9 ; 25 ; 26
35 ; 43 ; 100 ; 89 ; 12 ; 10

24.- Los siguientes son dos conjuntos de datos que corresponden a dos lotes de tablas que entran a una carpintería y en las cuales se consideró la variable x_i : número de defectos en cada tabla
Lote 1 : 10 ; 2 ; 3 ; 2 ; 4 ; 2 ; 5
Lote 2 : 50 ; 12 ; 13 ; 12 ; 14 ; 12 ; 15
a) Calcule la Mediana en cada lote
b) Compare los resultados y emita una conclusión.

25.- Examinando los registros de cuentas mensuales de una fábrica de muebles, el propietario solicita a su empleado tome los montos de las cuentas no pagadas durante el mes de mayo por sus clientes y le informe los resultados; los datos figuran a continuación:
$400; $ 1800; $1100; $700; $1000; $500; $3300; $900; $1200; $300; $1100; $1000; $ 600; $2600; $3700; $1500; $1800; $1000; $2100; $200
a) Calcule la Mediana y formule de acuerdo a ella el informe que le debe pasar el empleado al dueño.
b) Calcule el Modo.

26.- El siguiente cuadro representa las inasistencias diarias registradas en un mes en un vivero forestal:

X_i	0	1	2	3	4	5	9
f_i	1	2	7	6	2	1	1

Calcule la Mediana y el Modo.

27.- Los siguientes datos son diámetros (cm) medidos a 1.30m en

árboles de una plantación de álamos de 11 años de edad en el Dto. Capital (Pcia. de Sgo. del Estero):
19.4 ; 12.1 ; 16.8 ; 17.5 ; 18.7
Calcule la Media cuadrática.

28.- Calcule el Rango y Desvío Medio en las siguientes series simples:
5 ; 13 ; 9 ; 25 ; 26
35 ; 43 ; 100 ; 89 ; 12 ; 10

29.- Los siguientes son dos conjuntos de datos que corresponden a dos lotes de tablas que entran a una carpintería y en las cuales se consideró la variable x_i: número defectos en cada tabla
Lote 1 : 10 ; 2 ; 3 ; 2 ; 4 ; 2 ; 5
Lote 2 : 20 ; 12 ; 13 ; 12 ; 14 ; 12 ; 15
a) Calcule la Media Aritmética de cada lote, Desviación Estándar y Coeficiente de Variación.
b) Compare los resultados y emita una conclusión.

30.-El siguiente cuadro representa las inasistencias diarias registradas en un mes en un vivero forestal:

X_I	0	1	2	3	4	5	9
f_i	1	2	7	6	2	1	1

a) Calcule la Media Aritmética, Desviación Estándar, Variancia y Coeficiente de Variación.

31.- La siguiente es una tabla donde se muestra las alturas de álamos en una parcela de una plantación del Dto. Capital, Sgo. del Estero.

CLASES DE ALTURA (m) x_I	f_I
1 - 4	1
4 - 8	3
8 - 12	4
12 - 16	14
16 - 20	5
Total árboles	27

a) Calcule la altura promedio de la parcela.
b) Calcule Desviación Estándar, Variancia y Coeficiente de Variación.

32.- Defina probabilidad y diga entre qué valores varía.

33.- Se arroja un tetraedro con sus caras numeradas del 1 al 4. Describa el espacio muestral.

34.- Si el tetraedro es perfecto: a) ¿Cuál es la probabilidad de obtener el 4?, b) ¿Cuál es la probabilidad de obtener menos de 3 puntos?

35.-a) Escriba la distribución de probabilidades del número de puntos obtenidos al arrojar un tetraedro perfecto.
b) Represéntela gráficamente.

36.- Suponga que un tetraedro se arroja 1000 veces y las frecuencias obtenidas para cada cara figuran el la tabla que sigue:

xi (nº de puntos)	1	2	3	4	Total
f_i	200	180	320	300	1.000

a) Estime las probabilidades de cada cara del tetraedro .
b) ¿Que definición usó para calcularlas?

37.- De un mazo de 40 cartas españolas se extrae una al azar. ¿Cuál es la probabilidad de que: a) sea un oro o un basto? b) sea un as o una espada.

38.- La distribución uniforme de probabilidades es aquella que asigna a cada valor de la variable idéntica probabilidad. Dé un ejemplo y escriba la distribución elegida.

39.-Suponga que un viverista efectúa siembra directa en macetas, poniendo en cada maceta 3 semillas. Sabe que la probabilidad de que una semilla germine es 0.8. Considere la variable aleatoria x = nº de semillas que germinan en cada maceta. Escriba la distribución de probabilidades de la variable x suponiendo que sigue una distribución binomial.

40.- En la distribución del ejercicio anterior: a) ¿Cuál es la probabilidad de macetas con 2 plantas o menos? b) Si el viverista sembró 1000 macetas ¿cuántas se espera que estén vacías?

41.- En la distribución rectangular $X \sim R(0, 10)$:
a) Calcule la probabilidad de valores de x entre 5 y 8. Represente gráficamente a la probabilidad calculada.
b) Calcule la probabilidad de obtener valores mayores que 7. Represente gráficamente a la probabilidad calculada.
c) Para que una función sea considerada de densidad, además de no ser negativa, su integral en todo el espacio muestral debe ser igual a 1. Demuestre que en la distribución que nos ocupa, esta condición se cumple.

42.- Represente gráficamente a la distribución normal con media = 100 y desviación estándar = 10 y diga: a) donde se encuentra el máximo, b) donde están los puntos de inflexión.

43.- Dibuje en un mismo sistema de coordenadas dos distribuciones normales con distintas medias e igual variabilidad.

44.- Dibuje en un mismo sistema de coordenadas dos distribuciones normales con iguales medias y distinta variabilidad.

45.- El cambio de variable $Z = \frac{x - \mu}{\sigma}$ conduce a la normal estándar. a) ¿Qué parámetros definen a esta distribución?

46.- Para calcular probabilidad se dispone de dos tipos de tablas. Descríbalas.

47.- Utilizando la tabla de una cola resuelva el siguiente problema. Represente en cada caso la probabilidad buscada.
Un viverista tiene plantines de 1 año de edad cuyas alturas se distribuyen como una normal con media 50 cm y desviación estándar 5 cm. a) ¿Cuál es la probabilidad de que, al elegir una planta al azar, ésta tenga menos de 45 cm? b) ¿P(x > 45) ?
c) P(x < 58)? d) P(48<x<56) e) Un comprador le ofrece un buen precio pero con la condición de que los plantines tengan por lo menos 52 cm de altura. Si el viverista le envía, sin controlar, una remesa de 10000 plantines ¿cuántos serán rechazados?

48.- Las alturas de plántulas (en cm) de regeneración natural de un bosque se distribuyen normalmente con μ=11.5 y σ=2. Si se toma una planta al azar, cuál es la probabilidad de:

a) Qué sea menor a 10 cm?
b) Qué esté comprendida entre 9 y 12.5 cm?
c) Qué sea mayor a 13 cm?

49.- Los diámetros de los árboles de una plantación se distribuyen normalmente con media igual a 50 cm y desviación estándar 10 cm. Cuál es la probabilidad de encontrar:

a) árboles con diámetros menores a 57 cm?
b) árboles con diámetros mayores a 48 cm?
c) árboles comprendidos entre 45 y 55 cm?
d) entre qué valores se encuentra el 95% central de la población de diámetros?

49.- Usando la tabla de dos colas determine en la distribución del ejercicio anterior, los intervalos que corresponden al a) 68 % central , b) 95 % central y c) 99 % central y represéntelos gráficamente.

50.- En un monte, los quebrachos blancos de más de 30 cm tienen una altura de fuste media de 5 m y una desviación estándar de 1 m. Un acopiador recibe únicamente rollizos de 5 m como mínimo. Si la altura del tocón es generalmente 30 cm, ¿cuál es la probabilidad de que un rollizo sea rechazado?

51.-Un técnico debe seleccionar árboles semilleros en una plantación, y desea hacerlo de modo que queden aquellos ejemplares que constituyan el 5 % superior de diámetros. Si se sabe que los diámetros de esa plantación se distribuyen normalmente con media μ = 35 cm y desviación estándar σ = 3 cm, ¿cuál será el diámetro mínimo de los árboles seleccionados?

52.- Se ha determinado que la resistencia media a la compresión de la madera de una especie es 9 N/mm2 con desviación estándar de 2 N/mm2. a)¿Entre que valores se encuentra el 50 % de la población? b) La resistencia característica es aquel valor que deja por abajo el 5 % de los datos ¿cuánto vale aquí la resistencia característica?.

53.-Tome una muestra de 5 alumnos, anote su altura y utilice esa muestra para estimar por intervalo del 95 % de confianza a la altura media poblacional.

54.-Se desea saber la humedad final a la que se llega mediante un nuevo tratamiento de secado en madera de Juglans. Para ello se toma una muestra de n = 7 probetas que dan los siguientes resultados:

xi (H %) : 14.8, 17.5, 35.1, 43.2, 30.9, 31.4, 35.2.

a) Estime por intervalo de confianza del 95 % a la media poblacional μ de la humedad final a la que se llega con el nuevo método de secado.

b) ¿Cuantas probetas deberían ensayarse para que el error E no sea mayor de 4?

55.-La siguiente muestra corresponde a 10 parcelas de una plantación en cada una de las cuales se calculó el volumen en m3 por ha.

xi (vol. en m3/ha) : 350, 320, 311, 289, 365, 357, 288, 269, 396, 324.

a) Estime por intervalo de confianza del 99% al volumen medio poblacional.

b) ¿Cual es el error relativo?

c) ¿Cuál debe ser el tamaño de muestra para reducir el error relativo a la mitad?

56.-Se desea estimar la proporción de supervivientes en la regeneración natural de una especie. Para ello, en una parcela permanente, se cuentan las plantitas nacidas en un año dado (1293) y luego al año siguiente, las que sobrevivieron (753) . Estime por intervalo de confianza del 99 % a la P poblacional.

57.-En una fábrica de muebles se desea estimar la proporción de unidades con defectos leves de la producción. Para este fin, se eligen al azar 200 encontrándose 32 muebles con defectos leves. Determine el intervalo de estimación para una confianza del 90 %.

58.- En una inspección realizada a una plantación de árboles de 300 parcelas, al examinarla se encontró que el 25% de las parcelas estaban atacadas por insectos. El inspector debe confeccionar un informe donde conste la proporción poblacional de ataque en la plantación con una confianza del 99%. Determine esa proporción.

Bibliografía

Di Rienzo, J., Robledo, W., Guzmán, W., Balzarini, M., Casanoves, F., Gonzalez, L., Tablada, M. 2001. Infostat. Manual del usuario. Versión Estudiantil.

Peña Sánchez de Rivera, D. 1995. Estadística. Modelos y

Métodos. Fundamentos. Alianza Universidad textos. 565 pp.
Peña Sánchez de Rivera, D.(1995). Estadística Modelo y Métodos. Vol II:
Di Rienzo, J., Robledo, W., Balzarini M., Diaz, M., Gonzalez, L., Tablada, M., casanoves, F. 1995. Estadística para las Ciencias Agropecuarias.
Cuadras, Carles M.(1990). Problemas de Probabilidades y Estadística. Vol I: Probabilidades.411 p.
Cuadras, Carles M.(1990). Problemas de Probabilidades y Estadística. Vol II: Inferencia Estadística.452 p.
Cochran, W. 1974. Técnicas de muestreo. Cuarta Edición. Compañía Editorial Continental S.A. México. 507 p.
Freese, F. 1970. Métodos estadísticos elementales para técnicos forestales. Centro Regional de Ayuda Técnica. Agencia para el Desarrollo Internacional. México. 102 p.

.

Anexo

Arbol Nº	Dap (cm)	Ht (m)	nb	Observ.	Arbol Nº	Dap (cm)	Ht (m)	nb	Observ.
1	16.0	15.7	0	E	31	16.9		0	S
2	23.0	15.5	0	S	32	16.7		1	S
3	18.5	15.5	1	S	33	18.9		4	S
4	19.1	14.8	4	S	34	17.4		1	S
5	19.0	15.8	1	S	35	18.0		0	S
6	18.5	16.0	2	E	36	18.3		0	S

7	17.3	14.5	0	E	37	17.6		2	S
8	22.3	16.0	0	S	38	17.9		1	S
9	16.5	15.2	1	S	39	15.6		0	E
10	18.0	14.8	1	S	40	19.5	16.5	2	S
11	16.2	16.2	2	E	41				
12	17.4	16.8	3	S	42				
13	16.9	15.2	0	S	43				
14	16.3	15.8	1	S	44				
15	17.7	15.5	5	S	45				
16	12.7	14.5	4	E	46				
17	17.8	16.5	0	S	47				
18	18.8	16.5	0	S	48				
19	17.3	16.2	1	E	49				
20	17.4	14.8	1	S	50				
21	20.0	15.8	1	S	51				
22	15.2		2	E	52				
23	15.0		1	S	53				
24	18.3		1	S	54				
25	16.9		0	S	55				
26	16.4		0	S	56				
27	19.8	16.5	0	S	57				
28	15.8		0	S	58				
29	15.3		0	E	59				
30	17.6		1	S	60				