INGENIERÍA FORESTAL

Tecnológico de Estudios Superiores de Valle de Bravo

Manual de Prácticas

Nombre de asignatura: Estadística

Autor:

M.C. René García Martínez

Revisores:

M. EN EA. Felipe Neri Hernández Soto

ÍNDICE

PRÓLOGO

Las matemáticas son una herramienta primordial en el desempeño de actividades ingenieriles, ya que nos permiten describir fenómenos que de otro modo no pueden caracterizarse. La estadística, como rama de las matemáticas, se e8ncarga del estudio de la organización, análisis y presentación de datos, para la toma de decisiones.

Las actividades que realiza el Ingeniero Forestal comúnmente requieren de la recopilación, análisis y presentación de datos obtenidos en campo, por ello, dentro del plan de estudios que oferta la licenciatura de Ingeniería Forestal del Tecnológico de Estudios Superiores de Valle de Bravo (TESVB) se incluyen los cursos de Estadística, Muestreo y Regresión, y Diseños Experimentales.

Este manual está enfocado a cubrir los elementos básicos de la estadística descriptiva, como medidas de tendencia central, medidas de dispersión, métodos tabulares, medidas simultaneas de dos conjuntos de datos y estimación por intervalos. Finalmente se anexan fundamentos de Prueba de Hipótesis. La revisión de los conceptos implica en todo momento la recopilación previa de datos de campo, de tal manera que permita relacionar los conceptos teórico-prácticos.

I. Medidas de tendencia central

Cuando trabajamos con un conjunto de datos y evaluamos una variable, por ejemplo, la altura de los alumnos del semestre 401 del Tecnológico de Estudios Superiores de Valle de Bravo (TESVB) (Cuadro 1), nos percatamos que, al realizar la medición, no todas las personas evaluadas obtienen el mismo valor. Es notorio que algunos alumnos son más bajos y otros más altos, pero, una gran mayoría se acercan a un valor intermedio.

Cuadro 1. Altura de los alumnos del semestre 401 del TESVB

Alumno	Altura (m)	Alumno	Altura (m)
1	1.60	11	1.56
2	1.70	12	1.6
3	1.65	13	1.56
4	1.63	14	1.55
5	1.88	15	1.70
6	1.56	16	1.80
7	1.65	17	1.65
8	1.85	18	1.66
9	1.70	19	1.70
10	1.45	20	1.75

Las medidas de tendencia central hacen referencia al punto medio de una distribución de datos y también se denomina medidas de posición, se utilizan comúnmente los parámetros **media aritmética**, **moda y mediana**.

1.1 Media aritmética o promedio

En un grupo de datos, si deseamos conocer el punto medio, podemos calcular la media aritmética (promedio). Este parámetro estadístico se define como, la suma de los valores de cada una de las observaciones divididas entre el número total de datos. La fórmula matemática para calcularla es la siguiente:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

Dónde:

$\bar{y}$ = Media aritmética o promedio.
y_i = Observación de la variable en la posición i.
n = Número total de datos.

Para calcular el promedio del **Cuadro 1** se procede de la siguiente manera.

$$\bar{y} = \frac{y1 + y2 + y3 + y4 + y5 + y6 + \cdots + y20}{20}$$

$$\bar{y} = \frac{1.60 + 1.70 + 1.65 + 1.65 + 1.63 + 1.88 + \cdots + 1.75}{20}$$

$$\bar{y} = \frac{33.2}{20}$$

$$\bar{y} = 1.66\, m$$

1.1.1 Cálculo mediante Software Estadístico

La hoja de cálculo **Excel®** permite obtener la media aritmética con la función PROMEDIO. En el ejemplo del **Cuadro 1**, se calcula de la siguiente manera:

- Construir un cuadro de datos en la hoja de cálculo, introduciendo la variable con sus valores respectivos.

- Insertar la función **PROMEDIO** (Está función solicitará el conjunto de celdas que contenga los valores, a los cuales, se les quiere calcular la media aritmética) (Figura 1).

DESVEST.M | =PROMEDIO(B2:B21)

	A	B	C
1	Alumno	Altura (m)	
2	1	1.6	
3	2	1.7	
4	3	1.65	
5	4	1.63	
6	5	1.88	
7	6	1.56	
8	7	1.65	
9	8	1.85	
10	9	1.7	
11	10	1.45	
12	11	1.56	
13	12	1.6	
14	13	1.56	
15	14	1.55	
16	15	1.7	
17	16	1.8	
18	17	1.65	
19	18	1.66	
20	19	1.7	
21	20	1.75	
22	Media Aritmética	=PROMEDIO(B2:B21)	

A

I22

	A	B	C
1	Alumno	Altura (m)	
2	1	1.6	
3	2	1.7	
4	3	1.65	
5	4	1.63	
6	5	1.88	
7	6	1.56	
8	7	1.65	
9	8	1.85	
10	9	1.7	
11	10	1.45	
12	11	1.56	
13	12	1.6	
14	13	1.56	
15	14	1.55	
16	15	1.7	
17	16	1.8	
18	17	1.65	
19	18	1.66	
20	19	1.7	
21	20	1.75	
22	Media Aritmética	1.66	

B

Figura 1. A) Construcción de la tabla de datos e introducción de la función PROMEDIO y **B)** Resultado de la operación.

1.1.2 Ventajas y desventajas de la media

- Es uno de los parámetros estadísticos que cotidianamente utilizamos para analizar datos.
- Permite trabajar con datos cualitativos (color, atributos físicos, textura, sabor, otros) y cuantitativos (altura, peso, diámetro, temperatura, humedad, otros).
- Es útil para llevar a cabo procedimientos estadísticos como la comparación de medias de varios conjuntos de datos. En estadística inferencial es la medida de tendencia central que tiene mejores propiedades.
- Aunque la media es confiable en el sentido de que toma en cuenta todos los valores del conjunto de datos, puede verse afectada por valores extremos que no son representativos del resto de los datos. La media puede malinterpretarse si los datos no forman un conjunto homogéneo.

1.2 Mediana

Otra opción para calcular el punto medio de un grupo de datos es obteniendo el parámetro conocido como **Mediana (Me)**. Es el valor de la observación, que se encuentra a la mitad de un conjunto de elementos, cuando se ordenan de mayor a menor o de menor a mayor.

Dada la naturaleza de los datos, se pueden presentar 2 casos. El primero es que el total de elementos que conforman al conjunto de datos sea un número par. El segundo es que el total de elementos que conforman al conjunto de datos sea un número impar. Para ambos casos se desarrolla la metodología para su determinación.

1.2.1 Número Par

Para obtener la mediana, se calcula la media aritmética del par de datos que se encuentra a la mitad del conjunto, previamente ordenado.

$$Me = \frac{y_{\frac{n}{2}}}{2} + \frac{y_{\frac{n}{2}+1}}{2}$$

Dónde:

Me = Mediana

y = Valor numérico de la observación

n = número total de datos

Ordenando la variable **Altura** del grupo de alumnos del semestre 401 del TESVB podemos calcular la Me a partir del Cuadro 2.

Cuadro 2. Datos de la variable Altura, ordenados de manera ascendente.

Alumno	Altura (m)	Alumno	Altura (m)
1	1.88	11	1.65
2	1.85	12	1.65
3	1.80	13	1.63

4	1.75	14	1.60
5	1.70	15	1.60
6	1.70	16	1.56
7	1.70	17	1.56
8	1.70	18	1.56
9	1.66	19	1.55
10	1.65	20	1.45

La mediana se calcula de la siguiente manera

$$Me = \frac{y_{\frac{20}{2}}}{2} + \frac{y_{(\frac{20}{2})+1}}{2}$$

$$Me = \frac{y_{10}}{2} + \frac{y_{11}}{2}$$

$$Me = \frac{1.65}{2} + \frac{1.65}{2}$$

$$Me = 0.825 + 0.825$$

$$Me = 1.65$$

Nota: Recordar que el orden de los datos también puede hacerse de menor a mayor y el resultado debe ser el mismo.

1.2.2 Número impar

Cuando el total de elementos del conjunto es un número impar, la Mediana adquirirá el valor del dato que se encuentre a la mitad del grupo, previamente ordenado.

$$Me = y_{\frac{n+1}{2}}$$

Dónde:

Me = Mediana
y = Valor de la observación
n = Número total de datos

Para un estudio socioeconómico, el TESVB necesita realizar un análisis estadístico, para lo cual, cuantifica el número de alumnos que ingresaron a las distintas licenciaturas en el año 2014 (Cuadro 3), como medida de tendencia central se pide calcular la mediana.

Cuadro 3. Cantidad de alumnos de nuevo ingreso por licenciatura.

Número carrera	Licenciatura	Número de alumnos
1	Industrial	66
2	Administración	62
3	Gastronomía	60
4	Forestal	58
5	Eléctrica	51
6	Turismo	47
7	Arquitectura	40

La Me se calcula de la siguiente manera

$$Me = y_{\frac{n+1}{2}}$$

$$Me = y_{\frac{7+1}{2}}$$

$$Me = y_{\frac{8}{2}}$$

$$Me = y_4$$

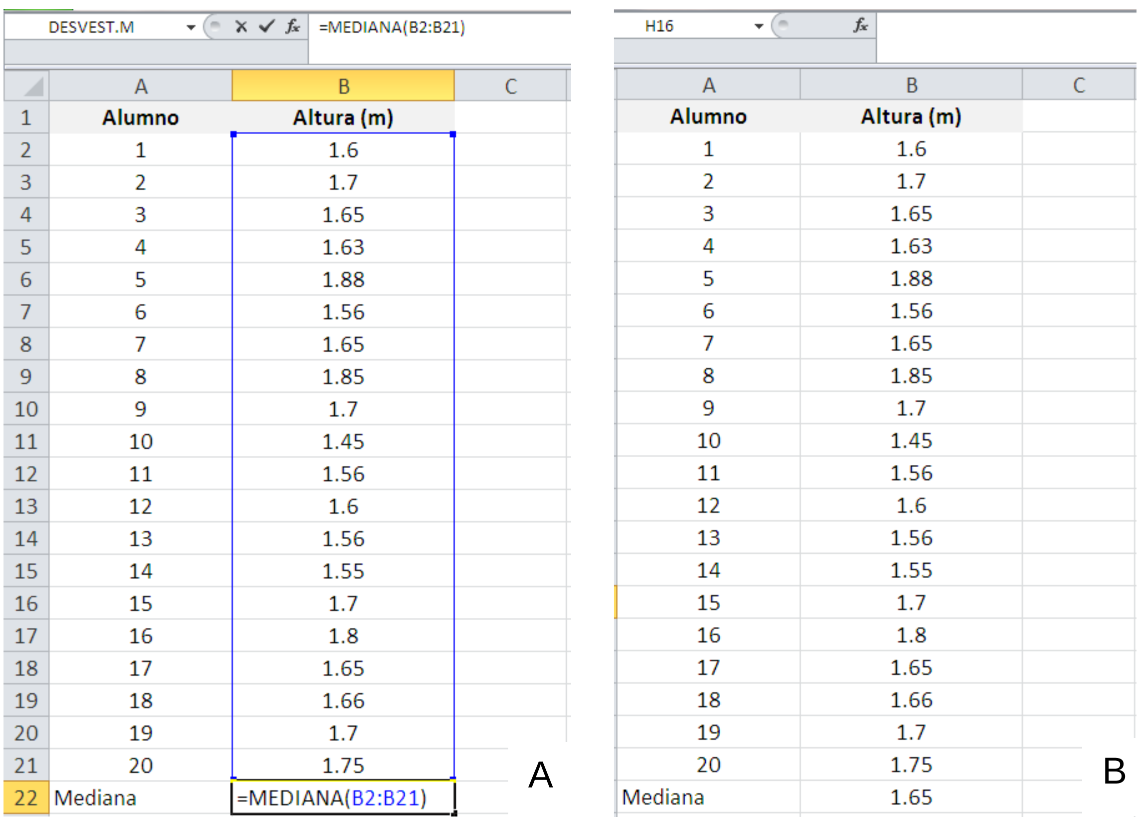
$$Me = 58$$

1.2.3 Cálculo mediante el uso del Software Estadístico

La hoja de cálculo de **Excel®** permite obtener la Me con la función MEDIANA. Para el caso del **Cuadro 1** se procede de la siguiente manera.

- Construir un cuadro de datos en la hoja de cálculo, introduciendo la variable con sus valores respectivos.

- Insertar la función **MEDIANA** (Está función solicitará el conjunto de celdas que contenga los valores, a los cuales, se les quiere obtener la mediana) (Figura 2).



DESVEST.M =MEDIANA(B2:B21)

	A	B	C
1	Alumno	Altura (m)	
2	1	1.6	
3	2	1.7	
4	3	1.65	
5	4	1.63	
6	5	1.88	
7	6	1.56	
8	7	1.65	
9	8	1.85	
10	9	1.7	
11	10	1.45	
12	11	1.56	
13	12	1.6	
14	13	1.56	
15	14	1.55	
16	15	1.7	
17	16	1.8	
18	17	1.65	
19	18	1.66	
20	19	1.7	
21	20	1.75	
22	Mediana	=MEDIANA(B2:B21)	

A

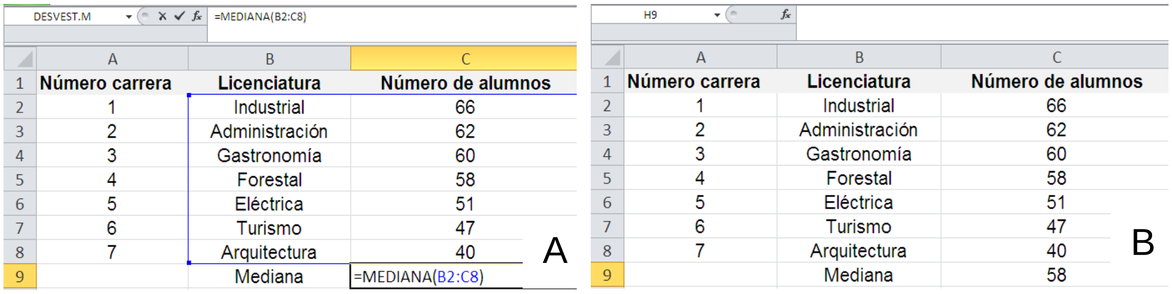
H16

A	B	C
Alumno	Altura (m)	
1	1.6	
2	1.7	
3	1.65	
4	1.63	
5	1.88	
6	1.56	
7	1.65	
8	1.85	
9	1.7	
10	1.45	
11	1.56	
12	1.6	
13	1.56	
14	1.55	
15	1.7	
16	1.8	
17	1.65	
18	1.66	
19	1.7	
20	1.75	
Mediana	1.65	

B

Figura 2. A) Construcción de la tabla de datos e introducción de la función MEDIANA y **B)** Resultado de la operación.

Para el ejemplo del **Cuadro 3** el procedimiento es el siguiente:



DESVEST.M =MEDIANA(B2:C8)

	A	B	C
1	Número carrera	Licenciatura	Número de alumnos
2	1	Industrial	66
3	2	Administración	62
4	3	Gastronomía	60
5	4	Forestal	58
6	5	Eléctrica	51
7	6	Turismo	47
8	7	Arquitectura	40
9		Mediana	=MEDIANA(B2:C8)

A

H9

	A	B	C
1	Número carrera	Licenciatura	Número de alumnos
2	1	Industrial	66
3	2	Administración	62
4	3	Gastronomía	60
5	4	Forestal	58
6	5	Eléctrica	51
7	6	Turismo	47
8	7	Arquitectura	40
9		Mediana	58

B

Figura 3. A) Construcción de la tabla de datos e introducción de la función MEDIANA y **B)** Resultado de la operación.

Nota: Observe que en la hoja de cálculo no es necesario ordenar los datos, ya que, al solicitar la función MEDIANA, el programa realiza todo el procedimiento.

1.2.4 Ventajas y desventajas de la Mediana

- Permite trabajar con datos cualitativos (color, atributos físicos, textura, sabor, otros) y cuantitativos (altura, peso, diámetro, temperatura, humedad, otros).
- No se ve afectada por los valores extremos. Esta es la propiedad más importante que tiene.
- Si hay un gran número de datos, el tener que ordenarlos para hallar la mediana insume esfuerzo y tiempo.

1.3 La moda

Finalmente, otra manera de expresar el punto medio de un grupo de elementos es mediante el uso de **la Moda (Mo).** Esta se define como el valor que se presenta con mayor frecuencia (más veces) dentro de un conjunto de datos.

El término Frecuencia se utiliza en varias áreas de las ciencias exactas. En estadística, hace referencia al número de veces que se repite algún elemento dentro de un grupo de datos cuantitativos o cualitativos.

Para el ejemplo del **Cuadro 1**, la Moda se calcula de la siguiente manera:

- Generar una gráfica o tabla de frecuencias
- La clase con el valor más alto en la frecuencia será la moda que se busca.

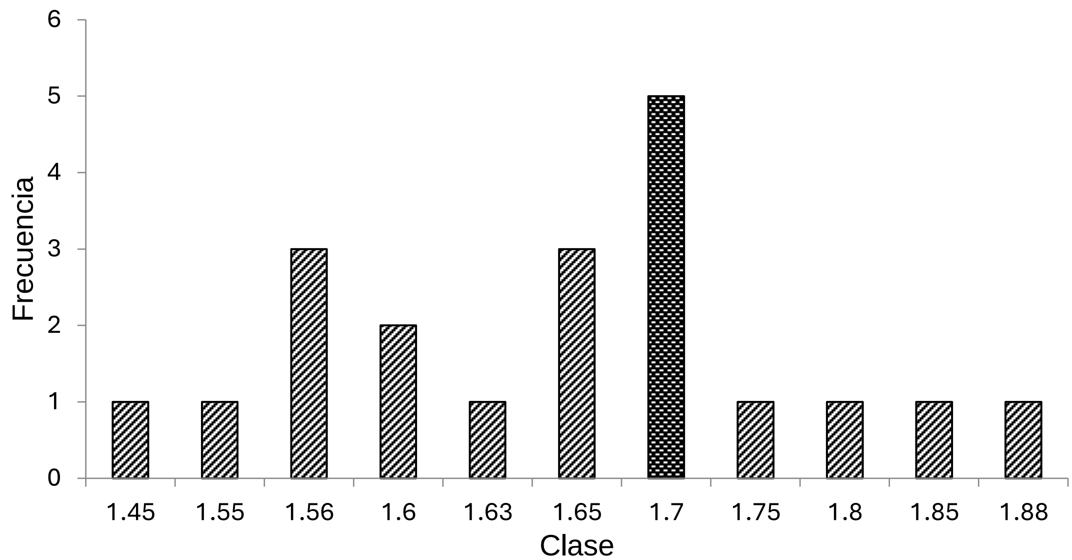


Figura 4. Gráfico de frecuencia para la altura de los alumnos de semestre 401 del TESVB.

En la Figura 4 se observa claramente que la barra de la clase 1.7 tiene el valor más grande en la frecuencia (5), por lo tanto, **la moda es 1.7.**

Cuadro 4. Distribución de frecuencias de las Alturas (Clases) de los alumnos del semestre 401 del TESVB.

Clase	Frecuencia
1.45	1
1.55	1
1.56	3
1.6	2
1.63	1
1.65	3
1.7	5
1.75	1
1.8	1
1.85	1
1.88	1

El **Cuadro 4** muestra que la clase 1.7 tiene el valor más alto para la frecuencia (5), por lo tanto, **la moda es 1.7.**

1.4.1 Cálculo mediante el uso del Software Estadístico

La hoja de cálculo de **Excel®** permite obtener la moda con la función **MODA.VARIOS**. Es importante aclarar que existe, también, la función **MODA.UNO**, la cual, solo determina un valor para la moda, pero existen casos en los cuales se presentan más de 2 valores que presentan la frecuencia más alta (multimodal) y no es posible obtenerlos mediante este procedimiento. El autor, recomienda utilizar la función **MODA.VARIOS**, puesto que, permite generar más de una moda en caso de que existiera.

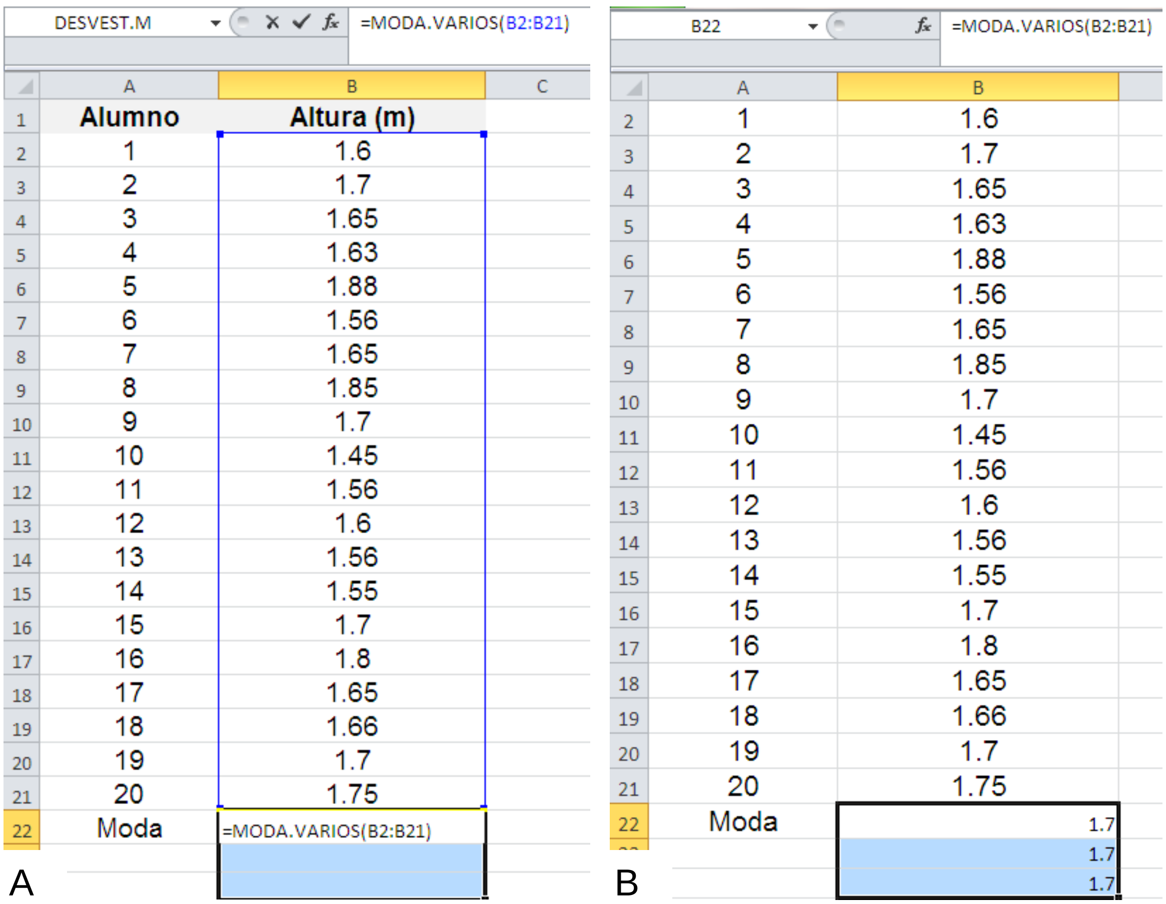
Para los datos del **Cuadro 1** se procede de la siguiente manera.

a) Construir una tabla de datos en la hoja de cálculo, introduciendo la variable con sus valores respectivos.

b) Seleccionar, debajo del último dato, las celdas en donde **Excel®** colocará los valores de la moda, el autor recomienda seleccionar mínimo tres celdas, insertar la función **MODA.VARIOS** (Está función solicitará el conjunto de celdas que contenga los valores, a los cuales, se les quiere obtener la moda) (**Figura 2**).

c) Una vez escrita la función MODA.VARIOS, junto con sus respectivas celdas, teclear el comando Ctrl+Mayus+Enter y el programa generará los valores de la moda.

d) En las celdas seleccionadas previamente, aparecerán los valores de la moda, si el valor es el mismo en todas las celdas, entonces, los datos presentan una sola moda. Si el valor es diferente en algunas celdas, entonces, los datos son multimodales. Dado que se pueden presentar varias modas, es necesario seleccionar una adecuada cantidad de celdas cuando se trabaje con grandes cantidades de datos.



DESVEST.M =MODA.VARIOS(B2:B21)

	A	B	C
1	Alumno	Altura (m)	
2	1	1.6	
3	2	1.7	
4	3	1.65	
5	4	1.63	
6	5	1.88	
7	6	1.56	
8	7	1.65	
9	8	1.85	
10	9	1.7	
11	10	1.45	
12	11	1.56	
13	12	1.6	
14	13	1.56	
15	14	1.55	
16	15	1.7	
17	16	1.8	
18	17	1.65	
19	18	1.66	
20	19	1.7	
21	20	1.75	
22	Moda	=MODA.VARIOS(B2:B21)	

A

B22 =MODA.VARIOS(B2:B21)

	A	B
2	1	1.6
3	2	1.7
4	3	1.65
5	4	1.63
6	5	1.88
7	6	1.56
8	7	1.65
9	8	1.85
10	9	1.7
11	10	1.45
12	11	1.56
13	12	1.6
14	13	1.56
15	14	1.55
16	15	1.7
17	16	1.8
18	17	1.65
19	18	1.66
20	19	1.7
21	20	1.75
22	Moda	1.7
		1.7
		1.7

B

Figura 4. A) Construcción de la tabla, selección de celdas e introducción de la función MODA.VARIOS y **B)** Resultado de la operación.

1.4.2 Ventajas y desventajas de la moda

- Permite trabajar con datos cualitativos (color, atributos físicos, textura, sabor, otros) y cuantitativos (altura, peso, diámetro, temperatura, humedad, otros).
- No resulta afectada por los valores extremos del grupo de datos.
- Cuando todas las clases de un grupo tienen la misma frecuencia, se dice que no tiene moda, por lo cual, no es muy factible en este caso.
- Cuando un conjunto de datos contiene 2 puntuaciones adyacentes con la misma frecuencia común (mayor que cualquier otra), la moda es el promedio de las 2 puntuaciones adyacentes. Por ejemplo, los datos (0,1,1,2,2,2,3,3,3,4,5) tienen una Mo = 2,5.

II. Medidas de dispersión

El simple hecho de calcular alguna medida de tendencia central no garantiza que nuestros datos sean precisamente confiables. Para, corroborar la precisión de un conjunto de números, además de la medida central, es necesario obtener los valores de los parámetros estadísticos conocidos como **medidas de dispersión.**

Las medidas de dispersión son útiles porque:

- Nos proporcionan información adicional que nos permite juzgar la confiabilidad de nuestra medida de tendencia central. Si los datos están muy dispersos, la posición central es menos representativa y visceversa.
- En muchas áreas de la ingeniería existen problemas característicos de distribuciones muy dispersas, por lo cual, debemos ser capaces de distinguir los grupos de datos que presentan alta dispersión antes de abordar los problemas.
- Nos permiten comparar varias muestras con promedios parecidos.

Ejemplos:

Los analistas financieros están preocupados por la dispersión de las ganancias de una empresa, ya que, los valores van desde muy grandes (positivos) hasta muy bajos negativos (negativos). Esto indica un riesgo mayor para los accionistas y para los acreedores.

De manera similar los expertos en control de calidad analizan los niveles de calidad de un producto. En la fabricación de galletas se requiere que el producto sea muy similar en cuanto a tamaño, peso y volumen, para ello, es necesario que los valores de las medidas de dispersión sean pequeños.

En una plantación comercial forestal, es necesario que los árboles tengan alturas y diámetros muy parecidos, esto se puede analizar cuando obtenemos las medidas de dispersión.

Las medidas de dispersión que se explicarán son el **Rango, la Varianza y la Desviación Estándar.**

2.1 El rango

En el ejemplo (Cuadro 1) de la altura de los alumnos del semestre 401 del TESVB, se presentan valores tanto muy altos como muy bajos que representan los extremos del grupo. Esta situación se presentará siempre que obtengamos datos de una variable de interés siempre y cuando se utilicen una gran cantidad de elementos de prueba.

Cotidianamente, cuando compramos un producto alimenticio, por ejemplo, naranjas, entre el grupo de frutos existe alguno que es más grande y otro que es más que pequeño que el resto.

El rango es una medida de dispersión que se pude definir como la diferencia entre el valor más grande y el valor más pequeño dentro de un grupo de elementos.

$$R = y_{mayor} - y_{menor}$$

Dónde:

R = Rango
Y mayor = Valor más bajo dentro del grupo de datos
Y menor = Valor más alto dentro del grupo de datos

Para el caso del **Cuadro 1**, el rango se calcula de la siguiente manera.

$$R = y_{>} - y_{<}$$
$$R = 1.88 - 1.45$$
$$R = 0.43$$

El rango, como medida de dispersión, no es difícil de entender y de calcular, pero su utilidad como medida de dispersión es limitada, ya que, sólo toma en cuenta el valor más alto y el valor más bajo e ignora la naturaleza de la variación entre todas las demás observaciones, además, tienen fuerte influencia los valores extremos. Debido a que considera sólo dos valores, tiene muchas posibilidades de cambiar drásticamente de una muestra a otra cuando se trabaja a partir de una población.

2.2 Varianza

La descripción más comprensible de la dispersión lo proporcionan aquellos parámetros que tratan con la desviación promedio con respecto a alguna medida de tendencia central. La Varianza puede calcularse para toda la población, aunque raramente trabajamos con la totalidad de los datos, pues comúnmente utilizamos muestras, es decir, una pequeña parte de la población.

Al elevar al cuadrado cada una de las distancias, logramos que todos los números que aparecen sean positivos y, al mismo tiempo asignamos más peso a las desviaciones más grandes. Las unidades de la varianza están elevadas al cuadrado (pesos al cuadrado, unidades al cuadrado, etc.) lo que hace que no sean claras o fáciles de interpretar.

2.2.1 Varianza Poblacional

La Varianza Poblacional se define como el promedio de las distancias al cuadrado que van de las observaciones a la media.

$$\sigma^2 = \frac{\sum_{i=1}^{N}({y_i}^2)}{N} - \mu^2$$ **Fórmula 1**

$$\sigma^2 = \frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N}$$ **Fórmula 2**

Dónde:

σ^2 = Varianza Poblacional.

N = Número total de datos en la población.

y_i = Valor de la variable "y" en la posición i.

μ = Media poblacional.

Tomando los datos del **Cuadro 1** como una población, la σ^2 se obtendría de la siguiente manera.

$$\sigma^2 = \frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N}$$

σ^2 = Desconocido.

N = 20.

μ = 1.66 m (se toma el valor de la media aritmética calculada previamente).

y = Toma los valores de la altura, de tal manera que (y_1 = 1.60, y_{20} = 1.75).

$$\sigma^2 = \frac{(y_1 - \mu)^2 + (y_2 - \mu)^2 + (y_3 - \mu)^2 + \cdots + (y_{20} - \mu)^2}{N}$$

$$\sigma^2 = \frac{(1.60 - 1.66)^2 + (1.70 - 1.66)^2 + (1.65 - 1.66)^2 + \cdots + (1.75 - 1.66)^2}{20}$$

$$\sigma^2 = \frac{0.0036 + 0.0016 + 0.0001 + \cdots + 0.0081}{20}$$

$$\sigma^2 = \frac{0.2132}{20}$$

$$\sigma^2 = 0.01066$$

En forma tabular, puede desarrollarse como lo muestra el **Cuadro 5**

Cuadro 5. Desarrollo del procedimiento para la Varianza Poblacional de manera tabular.

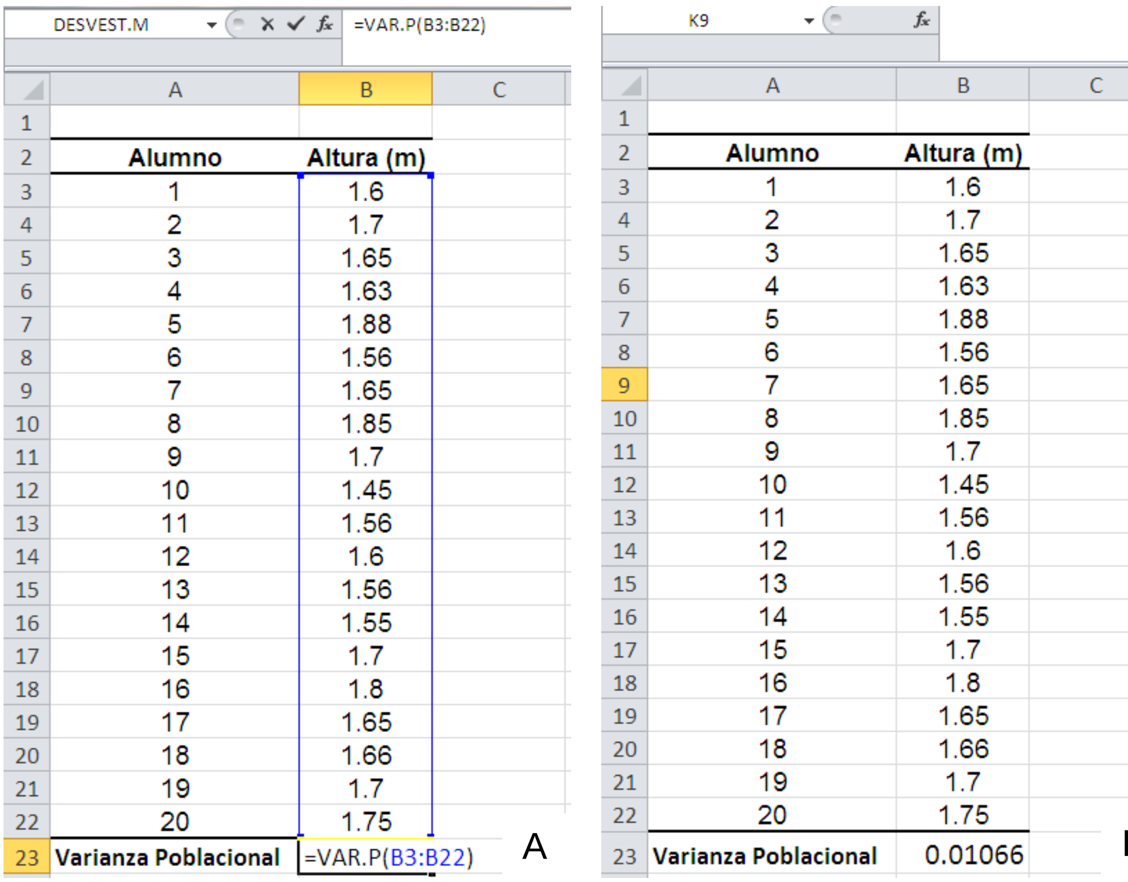
Alumno	Altura (m)	y_i-μ	$(y_i\text{-}\mu)^2$
1	1.60	1.60-1.66	0.0036
2	1.70	1.70-1.66	0.0016
3	1.65	1.65-1.66	0.0001
4	1.63	1.63-1.66	0.0009
5	1.88	1.88-1.66	0.0484
6	1.56	1.56-1.66	0.010
7	1.65	1.65-1.66	0.0001
8	1.85	1.85-1.66	0.0361
9	1.70	1.70-1.66	0.0016
10	1.45	1.45-1.66	0.0441
11	1.56	1.56-1.66	0.010
12	1.60	1.60-1.66	0.0036
13	1.56	1.56-1.66	0.010
14	1.55	1.55-1.66	0.0121
15	1.70	1.70-1.66	0.0016
16	1.80	1.80-1.66	0.0196
17	1.65	1.65-1.66	0.0001
18	1.66	1.66-1.66	0
19	1.70	1.70-1.66	0.0016
20	1.75	1.75-1.66	0.0081
		$\sum_{i=1}^{N}(y_i-\mu)^2$	0.2132
		σ^2	0.01066

Cálculo mediante software estadístico

La hoja de cálculo de **Excel®** permite obtener la Varianza Poblacional con la función VAR.P. Para el caso del **Cuadro 1** se procede de la siguiente manera.

- Construir un cuadro de datos en la hoja de cálculo, introduciendo la variable con sus valores respectivos.

- Insertar la función **VAR.P** (Está función solicitará el conjunto de celdas que contenga los valores, a los cuales, se les quiere calcular la varianza poblacional) (**Figura 5**).



DESVEST.M | =VAR.P(B3:B22)

	A	B	C
1			
2	Alumno	Altura (m)	
3	1	1.6	
4	2	1.7	
5	3	1.65	
6	4	1.63	
7	5	1.88	
8	6	1.56	
9	7	1.65	
10	8	1.85	
11	9	1.7	
12	10	1.45	
13	11	1.56	
14	12	1.6	
15	13	1.56	
16	14	1.55	
17	15	1.7	
18	16	1.8	
19	17	1.65	
20	18	1.66	
21	19	1.7	
22	20	1.75	
23	Varianza Poblacional	=VAR.P(B3:B22)	

A

K9

	A	B	C
1			
2	Alumno	Altura (m)	
3	1	1.6	
4	2	1.7	
5	3	1.65	
6	4	1.63	
7	5	1.88	
8	6	1.56	
9	7	1.65	
10	8	1.85	
11	9	1.7	
12	10	1.45	
13	11	1.56	
14	12	1.6	
15	13	1.56	
16	14	1.55	
17	15	1.7	
18	16	1.8	
19	17	1.65	
20	18	1.66	
21	19	1.7	
22	20	1.75	
23	Varianza Poblacional	0.01066	

B

Figura 5. A) Construcción de la tabla de datos e introducción de la función VAR.P y **B)** Resultado de la operación.

2.2.2 Varianza muestral (s^2)

Cuando se trabaja con una parte de la población, es decir, una muestra, entonces debe utilizarse la fórmula para la Varianza Muestral. Esta se define como la sumatoria de las distancias al cuadrado que van de las observaciones a la media dividido entre el número de datos menos uno.

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$ **Fórmula 1**

$$s^2 = \frac{\sum_{i=1}^{n}(y_i)^2}{n-1} - \frac{n.\bar{y}^2}{n-1}$$ **Fórmula 2**

s^2 = Varianza muestral.

n= Número total de datos en la muestra.

y_i = Valor de la variable "y" en la posición i.

$\bar{y}$ = Media muestral.

Tomando los datos del **Cuadro 1** como una muestra, la s^2 se obtendría de la siguiente manera.

$$s^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

s^2 = Desconocido

n = 20

$\bar{y}$ = 1.66 m (se toma el valor de la media aritmética calculada previamente).

y = Toma los valores de la altura, de tal manera que (y_1 = 1.60, y_{20} = 1.75).

$$s^2 = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + (y_3 - \bar{y})^2 + \cdots + (y_{20} - \bar{y})^2}{n-1}$$

$$s^2 = \frac{(1.60 - 1.66)^2 + (1.70 - 1.66)^2 + (1.65 - 1.66)^2 + \cdots + (1.75 - 1.66)^2}{20-1}$$

$$s^2 = \frac{0.0036 + 0.0016 + 0.0001 + \cdots + 0.0081}{20}$$

$$s^2 = \frac{0.2132}{19}$$

$$s^2 = 0.011221$$

En forma tabular puede desarrollarse como lo muestra el Cuadro 6.

Cuadro 6. Desarrollo del procedimiento para la Varianza Muestral de manera tabular.

Alumno	**Altura (m)**	**y_i-$\bar{y}$**	**$(y_i-\bar{y})^2$**
1	1.60	1.60-1.66	0.0036
2	1.70	1.70-1.66	0.0016
3	1.65	1.65-1.66	0.0001
4	1.63	1.63-1.66	0.0009
5	1.88	1.88-1.66	0.0484
6	1.56	1.56-1.66	0.010
7	1.65	1.65-1.66	0.0001
8	1.85	1.85-1.66	0.0361
9	1.70	1.70-1.66	0.0016
10	1.45	1.45-1.66	0.0441
11	1.56	1.56-1.66	0.010
12	1.60	1.60-1.66	0.0036
13	1.56	1.56-1.66	0.010
14	1.55	1.55-1.66	0.0121
15	1.70	1.70-1.66	0.0016
16	1.80	1.80-1.66	0.0196
17	1.65	1.65-1.66	0.0001
18	1.66	1.66-1.66	0
19	1.70	1.70-1.66	0.0016
20	1.75	1.75-1.66	0.0081
		$\sum_{i=1}^{N}(y_i-\bar{y})^2$	0.2132
		s^2	0.011221

Cálculo mediante software estadístico

La hoja de cálculo de **Excel®** permite obtener la Varianza Muestral con la función VAR.S. Para el caso del **Cuadro 1** se procede de la siguiente manera.

- Construir un cuadro de datos en la hoja de cálculo, introduciendo la variable con sus valores respectivos.
- Insertar la función **VAR.S** (Está función solicitará el conjunto de celdas que contenga los valores, a los cuales, se les quiere calcular la varianza poblacional) (**Figura 6**).

DESVEST.M | f_x =VAR.S(B2:B21)

	A	B
1	Alumno	Altura (m)
2	1	1.6
3	2	1.7
4	3	1.65
5	4	1.63
6	5	1.88
7	6	1.56
8	7	1.65
9	8	1.85
10	9	1.7
11	10	1.45
12	11	1.56
13	12	1.6
14	13	1.56
15	14	1.55
16	15	1.7
17	16	1.8
18	17	1.65
19	18	1.66
20	19	1.7
21	20	1.75
22	Varianza Muestral	=VAR.S(B2:B21)

A

K23 | f_x

	A	B
1	Alumno	Altura (m)
2	1	1.6
3	2	1.7
4	3	1.65
5	4	1.63
6	5	1.88
7	6	1.56
8	7	1.65
9	8	1.85
10	9	1.7
11	10	1.45
12	11	1.56
13	12	1.6
14	13	1.56
15	14	1.55
16	15	1.7
17	16	1.8
18	17	1.65
19	18	1.66
20	19	1.7
21	20	1.75
22	Varianza Muestral	0.011221053

B

Figura 6. A) Construcción de la tabla de datos e introducción de la función VAR.S y **B)** Resultado de la operación.

2.3 Desviación estándar

La desviación estándar dependerá directamente de la varianza. Se define matemáticamente como la raíz cuadrada de la varianza. Al existir Varianza Poblacional y Muestral, también se presentarán tanto Desviación Estándar Poblacional como Desviación Estándar Muestral.

2.3.1 Desviación Estándar Poblacional

Cuando se trabajan con datos de una población se utilizará alguna de las siguientes fórmulas.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}({y_i}^2)}{N} - \mu^2} \qquad \textbf{Fórmula 1}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \mu)^2}{N}} \qquad \textbf{Fórmula 2}$$

Dónde:

σ = Desviación estándar

σ^2 = Varianza Poblacional.

N = Número total de datos en la población.

y_i = Valor de la variable "y" en la posición i.

μ = Media poblacional.

Tomando los datos del **Cuadro 1** como una población, la σ se obtendría de la siguiente manera.

$$\sigma = \sqrt{\sigma^2}$$

σ = Desconocido

σ^2 = 0.01066

$$\sigma = \sqrt{0.01066}$$

$$\sigma = 0.1032$$

Cálculo mediante Software Estadístico

La hoja de cálculo **Excel®** permite obtener la media aritmética con la función PROMEDIO. En el ejemplo del **Cuadro 1**, se calcula de la siguiente manera:

- Construir un cuadro de datos en la hoja de cálculo, introduciendo la variable con sus valores respectivos.
- Insertar la función **DESVEST.P** (Está función solicitará el conjunto de celdas que contenga los valores, a los cuales, se les quiere calcular la Desviación Estándar Poblacional) (Figura 7).

DESVEST.M =DESVEST.P(B2:B21)

	A	B	C
1	Alumno	Altura (m)	
2	1	1.6	
3	2	1.7	
4	3	1.65	
5	4	1.63	
6	5	1.88	
7	6	1.56	
8	7	1.65	
9	8	1.85	
10	9	1.7	
11	10	1.45	
12	11	1.56	
13	12	1.6	
14	13	1.56	
15	14	1.55	
16	15	1.7	
17	16	1.8	
18	17	1.65	
19	18	1.66	
20	19	1.7	**A**
21	20	1.75	
22	Desviación Estandar	=DESVEST.P(B2:B21)	

A24

	A	B	C
1	Alumno	Altura (m)	
2	1	1.6	
3	2	1.7	
4	3	1.65	
5	4	1.63	
6	5	1.88	
7	6	1.56	
8	7	1.65	
9	8	1.85	
10	9	1.7	
11	10	1.45	
12	11	1.56	
13	12	1.6	
14	13	1.56	
15	14	1.55	
16	15	1.7	
17	16	1.8	
18	17	1.65	
19	18	1.66	
20	19	1.7	**B**
21	20	1.75	
22	Desviación Estandar	0.1032473	

Figura 7. A) Construcción de la tabla de datos e introducción de la función DESVEST.P y **B)** Resultado de la operación.

2.3.2 Desviación estándar muestral

En muy pocas ocasiones de trabaja con datos de la población total, lo más común es trabajar con una muestrea. En este último caso se utiliza alguna de las siguientes fórmulas.

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$$ **Fórmula 1**

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(y_i)^2}{n-1} - \frac{n.\bar{y}^2}{n-1}}$$ **Fórmula 2**

Tomando los datos del **Cuadro 1** como una población, la σ se obtendría de la siguiente manera.

$$s = \sqrt{s^2}$$

s = Desconocido

s^2 = 0.011221

$$s = \sqrt{0.011221}$$
$$s = 0.10592$$

Cálculo mediante Software Estadístico

La hoja de cálculo **Excel®** permite obtener la desviación estándar con la función PROMEDIO. En el ejemplo del **Cuadro 1**, se calcula de la siguiente manera:

- Construir un cuadro de datos en la hoja de cálculo, introduciendo la variable con sus valores respectivos.
- Insertar la función **DESVEST.M** (Está función solicitará el conjunto de celdas que contenga los valores, a los cuales, se les quiere calcular la Desviación Estándar Poblacional) (Figura 8).

DESVEST.M =DESVEST.M(B2:B21)

	A	B	C
1	Alumno	Altura (m)	
2	1	1.6	
3	2	1.7	
4	3	1.65	
5	4	1.63	
6	5	1.88	
7	6	1.56	
8	7	1.65	
9	8	1.85	
10	9	1.7	
11	10	1.45	
12	11	1.56	
13	12	1.6	
14	13	1.56	
15	14	1.55	
16	15	1.7	
17	16	1.8	
18	17	1.65	
19	18	1.66	
20	19	1.7	
21	20	1.75	
22	Desviación Estandar	=DESVEST.M(B2:B21	

A

A24

	A	B	C
1	Alumno	Altura (m)	
2	1	1.6	
3	2	1.7	
4	3	1.65	
5	4	1.63	
6	5	1.88	
7	6	1.56	
8	7	1.65	
9	8	1.85	
10	9	1.7	
11	10	1.45	
12	11	1.56	
13	12	1.6	
14	13	1.56	
15	14	1.55	
16	15	1.7	
17	16	1.8	
18	17	1.65	
19	18	1.66	
20	19	1.7	
21	20	1.75	
22	Desviación Estandar	0.1059295	

B

Figura 8. A) Construcción de la tabla de datos e introducción de la función DESVEST.M y **B)** Resultado de la operación.

Interpretación de la Desviación Estándar

La desviación estándar nos permite determinar, con un buen grado de precisión, dónde están localizados los valores de una distribución de frecuencias con relación a la media.

Para curvas cualesquiera, ***el teorema de Chebyshev*** asegura que al menos el 75% de los valores caen dentro de ± 2σ (2 desviaciones estándar) a partir de la media, y al menos el 89% de los valores caen dentro de ± 3σ.

Es posible puede medir con más precisión el porcentaje de observaciones que caen dentro de un rango específico de curvas simétricas con forma de campana (Figura 9):

- Aproximadamente 68% de las observaciones están dentro de ± 1s
- Aproximadamente 95% de las observaciones están dentro de ± 2s
- Aproximadamente 99% de las observaciones están dentro de ± 3s

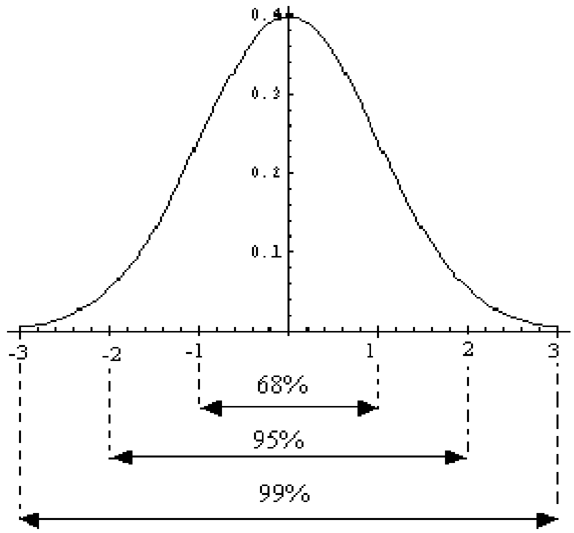


Figura 9. Representación de la Desviación Estándar con ayuda de la Campana *Gaussiana*.

En la Figura 9 interpretamos al 0 como **la media**, y los números como las unidades de **Desviación Estándar.**

Por ejemplo:

Interpretación	Para la Población	Para la Muestra
1 = Media + Desviación Estándar	$1 = \mu + \sigma$	$1 = \bar{Y} + s$
-1 = Media – Desviación Estándar	$-1 = \mu - \sigma$	$-1 = \bar{Y} - s$
2 = Media + 2Desviación Estándar	$2 = \mu + 2\sigma$	$2 = \bar{Y} + 2s$
-2 = Media – 2Desviación Estándar	$-2 = \mu - 2\sigma$	$-2 = \bar{Y} - 2s$

2.4 Coeficiente de variación

Este parámetro es una medida dispersión relativa de un conjunto de datos, que se obtiene dividiendo la desviación estándar del conjunto de datos entre su media aritmética, y se expresa en porcentaje. Se calcula mediante la siguiente fórmula.

$$\boldsymbol{C.V.} = \frac{S}{Media\ Aritmética}\ x\ 100$$

Dónde:

C.V. = Coeficiente de variación.
S = Desviación estándar.

Para el caso del **Cuadro 1** se procede de la siguiente manera;

$$\boldsymbol{C.V.} = \frac{0.1059295}{1.66}\ x\ 100$$

C.V. = 6.38 %

2.4.1 Interpretación del coeficiente de variación

Es independiente de la unidad de medida de los datos, por esta razón es útil para comparar conjuntos de datos donde las escalas de medición son diferentes. El problema de este parámetro estadístico es que deja de ser útil cuando la media aritmética es cercana a cero.

2.5 Desviación media

La desviación media (D.M.) se define como el promedio de las desviaciones de los datos de un conjunto con respecto a la media, para ello se utiliza la siguiente fórmula:

$$\textbf{D.M.} = \frac{\sum_{i=1}^{n} |x_i - promedio|}{\text{n}}$$

Ejemplo:

En una muestra de árboles del vivero de la Protectora de Bosques del Estado de México (PROBOSQUE) se obtuvieron los siguientes resultados en altura (cm): 12, 14, 10, 15, 11, 12. Con estos datos se calculó la desviación media (Cuadro 7).

Cuadro 7. Cálculo de la desviación media de los árboles de *Pinus pringlei* de 4 meses de edad.

Árbol	Altura (cm)	X_i-Media	$\lvert X_i$-Media$\rvert$
1	12	-0.4	0.4
2	14	1.6	1.6
3	10	-2.4	2.4
4	15	2.6	2.6
5	11	-1.4	1.4
Promedio	12.4	**D.M.**	1.68

2.5.1 Cálculo mediante software estadística

La hoja de cálculo **Excel®** permite obtener la desviación media utilizando la función DESVPROM. Para el caso del Cuadro 7 se procede de la siguiente manera.

- Construir un cuadro de datos en la hoja de cálculo, introduciendo la variable con sus valores respectivos.
- Insertar la función **DESVPROM** (Está función solicitará el conjunto de celdas que contenga los valores, a los cuales, se les quiere calcular la desviación promedio) (**Figura 10**).

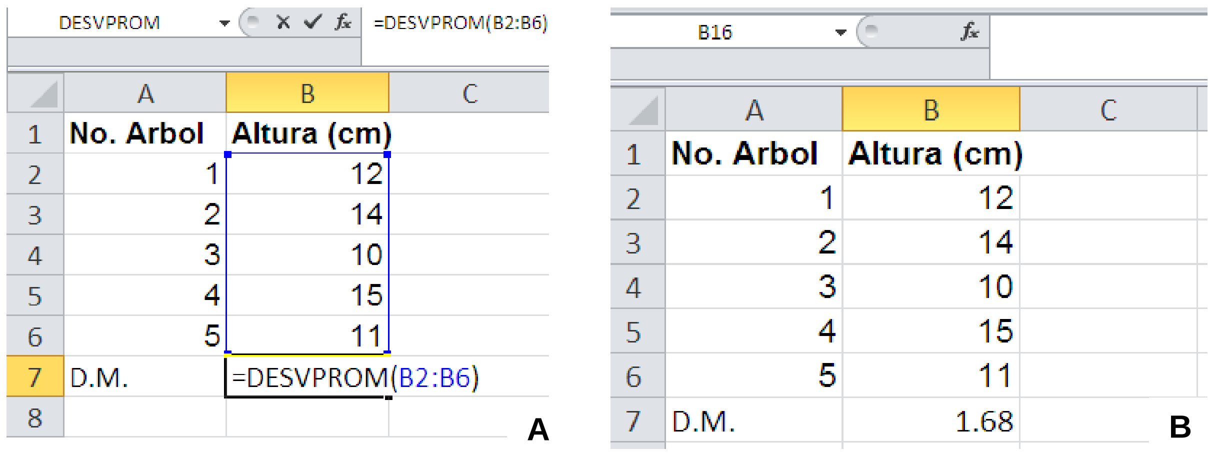


Figura 10. A) Construcción de la tabla de datos e introducción de la función DESVPROM y **B)** Resultado de la operación.

III. Métodos tabulares

3.1 Conceptualización

Distribución de frecuencias: Es la ordenación de un conjunto de datos dentro de una tabla que se divide en diferentes clases o categorías, registrando el número de individuos u objetos que perteneces a cada clase con el fin de facilitar su análisis e interpretación.

Clases o categorías: Número de grupos en que se divide del conjunto de datos a analizar. El número de clases que tendrá una tabla de frecuencias simples se determina en forma arbitraria por el investigador, aunque se recomienda elegir un número de clases entre 5 y 10. Dependiendo del número de clases elegidas se tendrán los valores de los intervalos y de sus anchuras.

Intervalos de Clase: El valor mínimo y el máximo para los datos que conforman una clase específica.

Frecuencia de Clase. Número de individuos u objetos que pertenecen a una clase específica.

Límite inferior de clase (L.I.): Es el valor mínimo para los datos que deben conformar una clase específica.

Límite superior de Clase (L.S.): Es el valor máximo para los datos que deben conformar una clase específica.

Anchura del intervalo de clase: Es la cantidad de unidades o fracciones de unidad que están contenidas en un intervalo de clase específico. Se obtiene a partir del límite superior de clase (L.S.) restando el límite inferior de clase (L.I.) y al resultado de la resta sumando la diferencia entre valores adyacentes en la clase. La diferencia entre los valores adyacentes en una clase específica es conocida como la escala de aumento de los datos en un intervalo. **Anchura =** L.S. – L.I. + Escala de Aumento.

Marca de clase: Es el punto medio de una clase específica y se obtiene sumando el límite inferior y superior de una clase y dividiendo la suma entre 2. **Marca de Clase =** (L. S. + L. I.)/2

Ejemplo:

En una plantación comercial de *Abies religiosa* de 2 años de edad se realizó un análisis de la situación actual de los árboles en cuanto a la altura, la unidad de medida fueron metros. La información se presenta en el Cuadro 8.

Cuadro 8. Tabla de distribución de frecuencia simple de los árboles de una plantación de *Abies religiosa*.

Altura	Número de Árboles
1.0 – 1.9 (1)	60
2.0 – 2.9 (2)	140
3.0 – 3.9 (3)	120
4.0 – 4.9 (4)	90

Número de Clases: 4 Clases (1.0-1.9, 2.0-2.9, 3.0-3.9 y 4.0 – 4.9).

Frecuencia de clases: En la Clase 1 se contaron 60 árboles, en la Clase 2 (140 árboles), en la clase 3 (120 árboles) y en la clase 4 (90 árboles).

Límite inferior y superior de clase: en la clase 1 (L.I. = 1.0 y L.S. = 1.9), en la Clase 2 (L.I.= 2.0 y L.S. = 2.9), en la Clase 3 (L.I.=3.0 y L.S.= 3.9) y en la Clase 4 (4.0 y 4.9).

Anchura del Intervalo de Clase = L.S. – L.I. + Escala de Aumento. (Tomando como ejemplo la Clase 1). **Anchura =** 1.9 - 1.0 + 0.1 = 1.0.

Marca de clase = (1.9 + 1.0)/2 = 1.45.

3.2 Tipos de distribución de frecuencia

Distribución o tablas de frecuencias simples: Es la ordenación de un conjunto de datos dentro de una tabla que es dividida en diferentes clases con intervalos cerrados, registrando el número de individuos u objetos que pertenecen a cada clase.

Distribución o tabla de frecuencias simples: Es la ordenación de un conjunto de datos dentro de una tabla que es dividida en diferentes clases con intervalos cerrados, registrando el número de individuos u objetos que pertenecen a cada clase y expresando este número como un porcentaje relativo con respecto a la suma total de las frecuencias en la tabla. Un intervalo de clase que presenta sus límites superior e inferior bien definidos se conoce como intervalo de clase cerrado.

Distribución o tabla de frecuencias acumuladas: Es la ordenación de un conjunto de datos dentro de una tabla de frecuencias en diferentes clases con intervalos abiertos, registrando el número de individuos u objetos que pertenecen a cada clase. Los límites de los intervalos abiertos y el número de individuos en cada clase se obtienen tomando como referencia la tabla de frecuencias simples. Para obtener la frecuencia de una clase específica se suman las frecuencias de la clase equivalente y de las clases que le anteceden o preceden en la tabla de frecuencias simples. Un intervalo de clase que, aparentemente, no tiene límite superior o inferior, se le conoce como intervalo de clase abierto.

Distribución o tabla de frecuencias relativas acumuladas: Es la ordenación de un conjunto de datos dentro de una tabla que es dividida en diferentes clases con intervalos abiertos, registrando el número de individuos u objetos que pertenecen a cada clase y expresando este número como un porcentaje relativo con respecto a la suma total de las frecuencias en la tabla. Los límites de los intervalos abiertos y el porcentaje relativo en cada clase se obtienen tomando como referencia la tabla de frecuencias relativas. Para obtener el porcentaje de una clase específica se suman los porcentajes relativos de la clase equivalente y de las clases que le anteceden o preceden en la tabla de frecuencias relativas.

Ejemplo:

Cuadro 9: Tabla de frecuencia simple y frecuencia relativa de la altura de los árboles de un plantación de *Abies religiosa*.

Altura	Número de Árboles	Frecuencia Relativa
1.0 – 1.9 (1)	60	14.634
2.0 – 2.9 (2)	140	34.146
3.0 – 3.9 (3)	120	29.268
4.0 – 4.9 (4)	90	21.951

Cuadro 10: Tabla de frecuencia simple y frecuencia acumulada de la altura de los árboles de una plantación de *Abies religiosa*.

Altura	Frecuencia Simple	Altura	Frecuencia Acumulada
1.0 – 1.9 (1)	60	Menos de 2.0 (1)	60
2.0 – 2.9 (2)	140	Menos de 3.0 (2)	200
3.0 – 3.9 (3)	120	Menos de 4.0 (3)	320
4.0 – 4.9 (4)	90	Menos de 5.0 (4)	410

Cuadro 11. Tablas de frecuencia simple, acumulada, relativa y relativa acumulada.

Altura	F. Simple	F. Acumulada	F. Relativa	F. Relativa Acumulada
1.0 – 1.9 (1)	60	60	14.634	14.634
2.0 – 2.9 (2)	140	200	34.146	48.780
3.0 – 3.9 (3)	120	320	29.268	78.049
4.0 – 4.9 (4)	90	410	21.951	100.000

3.3 Gráficas de la distribución de frecuencias

Existen tres tipos de representaciones gráficas: Los histogramas, los polígonos de frecuencias y las ojivas.

Histograma: Representación gráfica por medio de barras de una distribución de frecuencias simples o relativas.

Polígonos de frecuencias: Representación gráfica por medio de líneas de una distribución de frecuencias simples o relativas.

Ojiva: Representación gráfica por medio de líneas de una distribución de frecuencias acumuladas o relativas acumuladas.

Ejemplos:

Histograma: Para el histograma de frecuencias simples se calculan en primer lugar las marcas de clases de los diversos intervalos y se representan en el eje X. En el eje Y se presenta el número de conjuntos de individuos que se encuentran dentro de los diferentes intervalos (Figura 11).

Para el histograma de frecuencias relativas se vuelven a utilizar las marcas de clases de los diversos intervalos y se representan en el eje X. En eje Y se representan las frecuencias relativas de los diferentes intervalos (Figura 12).

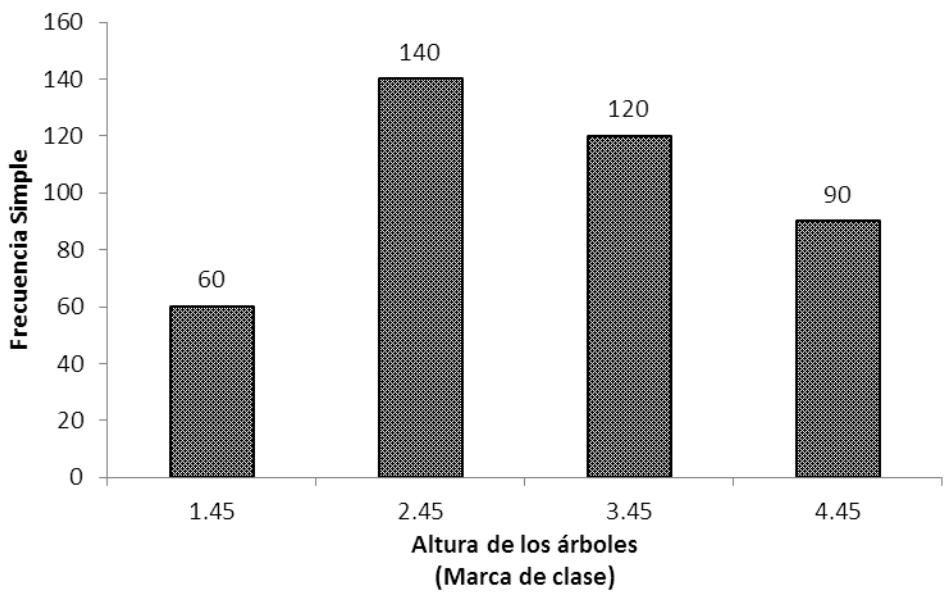


Figura 11: Histograma de las distribución de frecuencias simples de la altura de los árboles de la plantación de *Abies religiosa.*

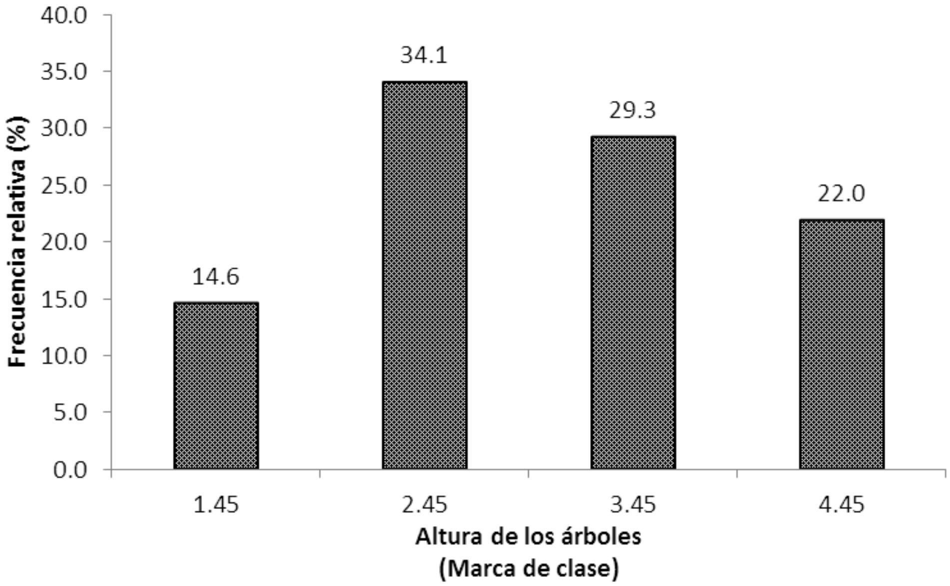


Figura 12: Histograma de la distribución de frecuencias simples de la altura de los árboles de la plantación de *Abies religiosa.*

Polígono de frecuencias: Consiste en una serie de coordenadas rectangulares, unidas por una línea, que representan las frecuencias de los diversos intervalos que constituyen una tabla de frecuencias simples o relativas. En el eje X se representan las frecuencias simples o relativas de los diferentes intervalos. Pueden realizarse polígono de frecuencias simples (Figura 13) y relativas (Figura 14).

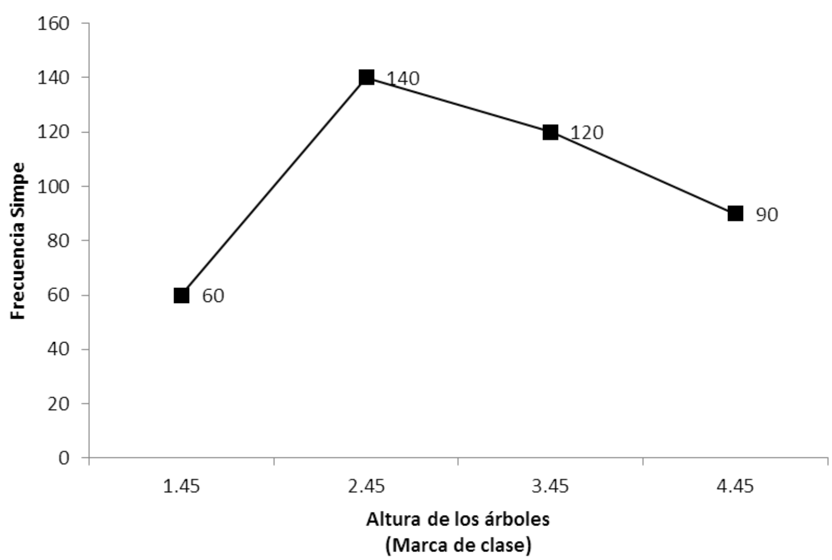


Figura 13. Polígono de la distribución de frecuencias relativas de la altura de los árboles de la plantación de *Abies religiosa.*

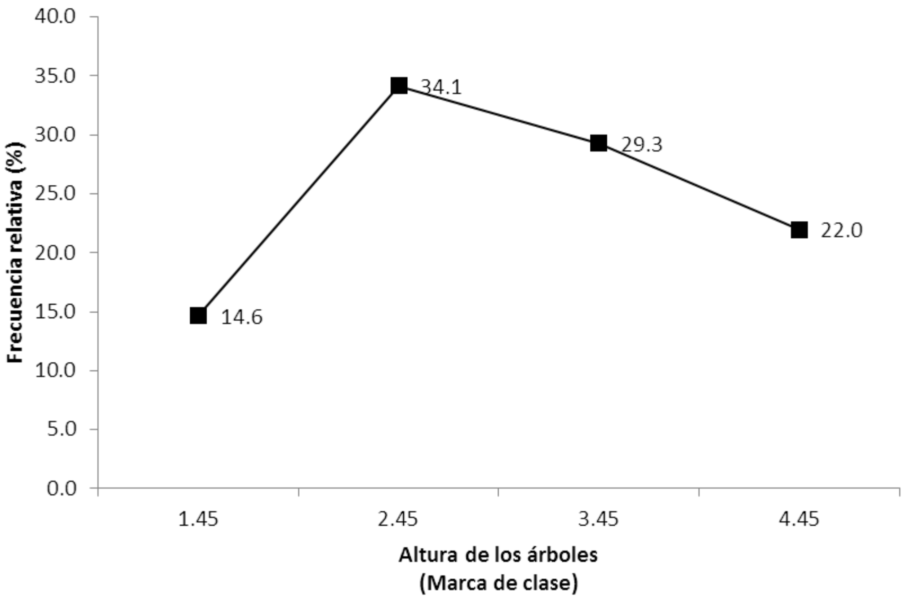


Figura 14. Polígono de la distribución de frecuencias relativas de la altura de los árboles de la plantación de *Abies religiosa.*

Ojiva de frecuencias: Consiste en una serie de coordenadas rectangulares unidas mediante líneas que representan las frecuencias de los diversos intervalos que constituyen una tabla de frecuencias acumuladas o relativas acumuladas. En el eje X se representan los intervalos mediante su límite superior o inferior. En el eje Y se representan las frecuencias acumuladas (Figura 15) o relativas acumuladas (Figura 16) de los distintos intervalos.

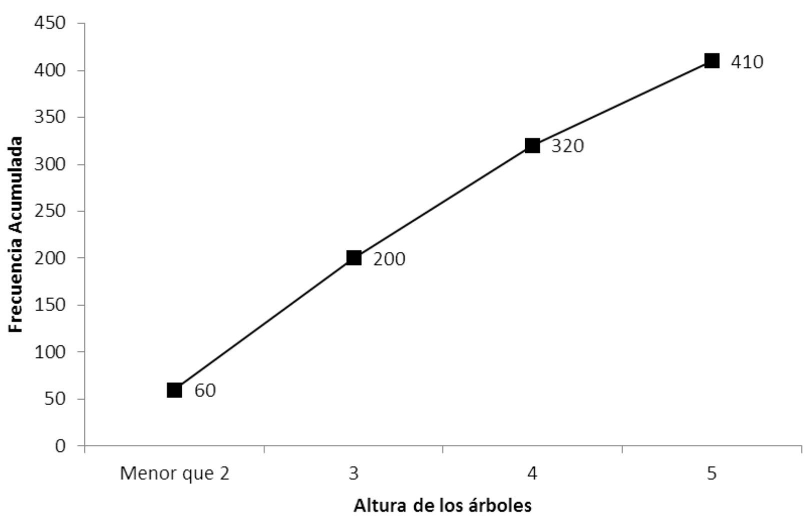


Figura 15. Ojiva de frecuencia acumulada de la altura de los árboles en la plantación de *Abies religiosa.*

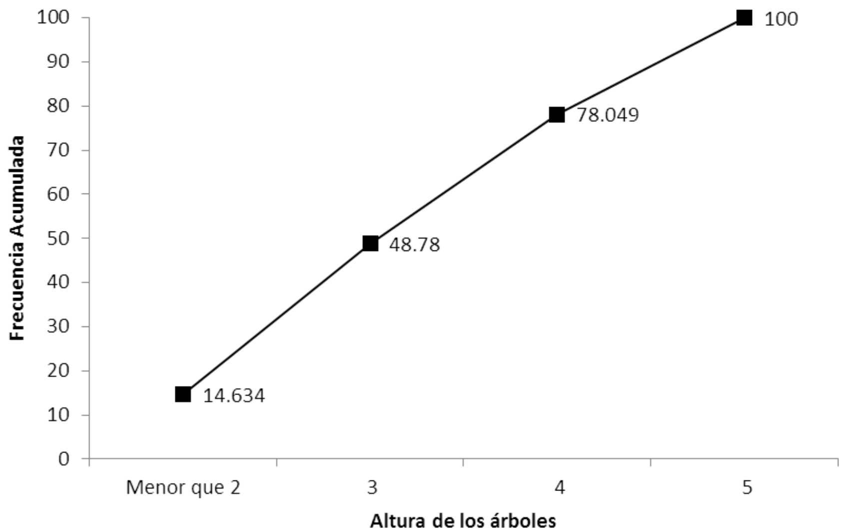


Figura 16. Ojiva de frecuencia relativa acumulada de la altura de los árboles en la plantación de *Abies religiosa.*

IV. Prueba de hipótesis

Las ciencias forestales continuamente requieren del análisis estadístico de datos obtenidos en campo. Para ello hacemos uso de técnicas que nos ayudan a tomar decisiones acerca de un fenómeno de estudio.

4.1 Conceptualización

En los Diseños Experimentales se utiliza el procedimiento conocido como Prueba de Hipótesis. La hipótesis, en la investigación científica, es un elemento muy importante ya que es la teoría que el investigador pone a prueba durante la ejecución de su proyecto.

Pájaro-Huertas (2002) menciona varias definiciones acerca de la formulación de hipótesis (Cuadro 1). Estas nos dan un panorama muy amplio de su significado y la forma en que debe abordarse al momento de plantear un experimento.

Prueba de Hipótesis: Método estadístico que se emplea para determinar si una hipótesis es verdadera o falsa (Castillo-Márquez, 2006).

De acuerdo con Castillo-Márquez (2006) existen elementos indispensables que debe contener una prueba de hipótesis y son las siguientes:

Hipótesis a Probar: Consiste en un par de hipótesis (*Hipótesis Nula e Hipótesis Alternativa*) las cuales se contraponen, es decir la Hipótesis Alternativa es contraria a la Hipótesis Nula. La Hipótesis Nula es aquella que el investigador está dispuesto a sostener como cierta y se representa como Ho. La Hipótesis Alternativa es aquella que contradice a la Hipótesis Nula y se representa como Ha.

Estadística de Prueba: Es una fórmula estadística que, con base en los datos experimentales, permite obtener un número (Valor calculado) que se compara contra un valor de tablas (Valor tabulado) de la distribución de probabilidad con la que se relaciona la estadística de prueba. Existen varias estadísticas de prueba que se utilizan en diseños experimentales y la más común es el Análisis de la Varianza (ANOVA).

Regla de decisión: Determina la forma en debe relacionarse el valor calculado con el valor tabulado para rechazar o no rechazar la Ho.

Conclusión: Después de rechazar o no rechazar la Ho se debe redactar la conclusión pertinente con base en el experimento.

Cuadro 12. Diversos enfoques en la conceptualización de la hipótesis.

No.	Concepto
1	**Etimológicamente (Hipo = Bajo) (Thesis = Posición y situación):** Explicación supuesta que está bajos ciertos criterios hechos, a los que sirve de soporte.
2	Es la suposición que permite establecer una relación entre hechos.
3	Es una afirmación sujeta a confirmación.
4	Es una solución teórica o tentativa a un problema.
5	Es una explicación provisional de un problema.
6	Es una relación entre dos o más variables para describir o explicar un problema.
7	Es un raciocinio o una conclusión según la cual un determinado conjunto de fenómenos, cuyo pensamiento forma un predicado del juicio, puede ser explicado como el resultado de un orden sujeto a leyes que se observa directamente.
8	Es el juicio problemático mediatizado sobre el vínculo sujeto a leyes de los fenómenos, que se obtiene como deducción de un raciocinio de probabilidad.
9	Es una suposición acerca de la existencia de una entidad, la cual permite la explicación de los fenómenos o del fenómeno estudiado.
10	Es aquella formulación que se apoya de un sistema de conocimientos organizados y sistematizados, y que establece una relación entre dos o más variables para explicar y predecir en la medida del o posible, aquellos fenómenos de una parcela determinada de la realidad en caso de comprobarse relación establecida.
11	Conjunto de datos que describen un problema, donde se propone una reflexión y/o explicación que plantea la solución a este problema.

12	Enunciado o proposición que sirve de antecedentes para explicar ¿Por qué? o ¿Cómo? Se produce un fenómeno o conjunto de fenómenos relacionados entre sí.

Fuente: Pájaro- Huertas, 2002.

4.2 Diseño Experimental Completamente Aleatorizado

En las ciencias forestales se ha utilizado la experimentación para evaluar el efecto de tratamientos sobre alguna variable de interés. Por ejemplo, se ha comparado la influencia del pastoreo de cabras sobre la cantidad de arbustos que crecen en un pastizal, con la finalidad de eliminar fuente de combustión para prevenir incendios forestales. También se han probado distintas concentraciones de Deltametrina para evaluar el porcentaje de insectos descortezadores (*Dendroctonus adjunctus*) muertos después de la aplicación en campo. Otro ejemplo es la influencia que tiene el aclareo de áreas dentro de un bosque para conocer la cantidad de renuevo establecido.

En todos estos casos se puede utilizar un diseño experimental completamente aleatorizado (DECA) ya que es el más sencillo, eficiente y se origina por la asignación aleatoria de los tratamientos a un conjunto de unidades experimentales previamente determinado (Badii *et al*., 2007).

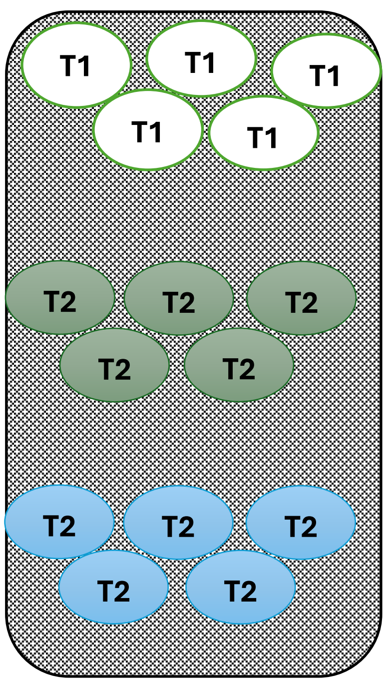
Se emplea cuando las unidades experimentales son básicamente homogéneas entre sí, es decir cuando la variación entre ellas es pequeña. También es utilizado preferentemente en experimentos de laboratorio o invernadero porque es posible, en estos casos, mantener las condiciones homogéneas en las unidades experimentales.

4.2.1 Proceso de aleatorización

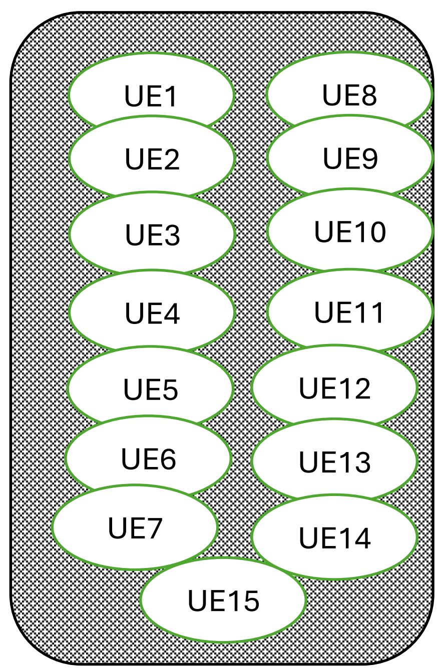
El proceso de aleatorización requiere de unidades experimentales homogéneas, las cuales deberán identificarse con un código. Los tratamientos también deben identificarse y designar cuantas repeticiones tendrá cada uno. En dos urnas se colocan por un lado los tratamientos y en el otro las unidades experimentales. El proceso de selección comienza seleccionando un tratamiento y una U.E. y se forma la primera combinación. El proceso continua hasta agotar tanto tratamientos y U.E.

Ejemplo.

Mediante un Diseño Experimental Completamente Aleatorizado se compararon tres tratamientos de fertilización (T1 = 0N-0P-0K, T2= 10N-10P-10K, T3 = 20N-20P-20K) que se aplicaron a árboles de roble. Cada tratamiento incluyó 5 repeticiones. Se aplicó la técnica de aleatorización quedando la distribución de los tratamientos como se muestra en la figura.



Urna Tratamientos



Urna UE

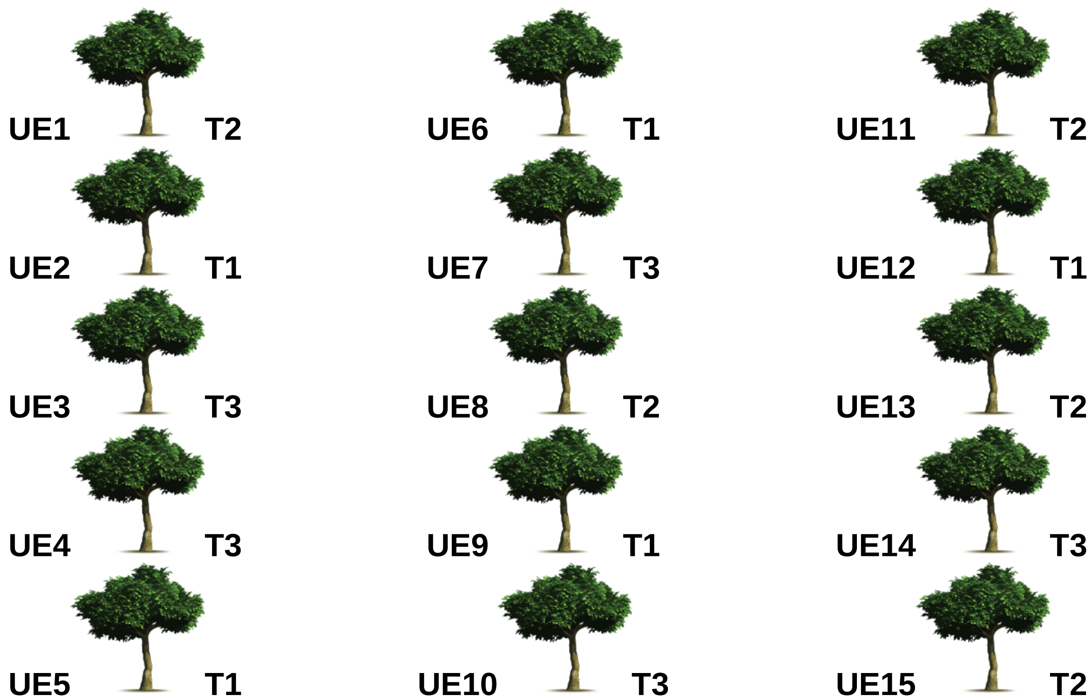


Figura 17. Proceso de aleatorización a través del cual se asignan los tratamientos a las unidades experimentales.

Cuadro 13. Fórmulas para el Análisis de la Varianza de un Diseño Experimental Completamente Aleatorizado.

FV	GL	SC	CM	F Cal	F tab	Sig.
Trat	t-1	$(\sum_{i=1}^{t} \frac{y_{i\bullet}{}^2}{r}) - \frac{y_{\bullet\bullet}{}^2}{tr}$	$\frac{\mathrm{SC(TRA)}}{\mathrm{GL(TRA)}}$	$\frac{\mathrm{CM(TRA)}}{CM(ERROR)}$	$F_{GL\ (Error)}{}^{GL\ (Tra)}$	* ** NS
Error	gl(tot)-gl(trat)	SC(tot)-SC(trat)	$\frac{\mathrm{SC(Error)}}{\mathrm{GL(Error)}}$			
Total	tr-1	$\sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij}{}^2 - \frac{y^2\cdot\cdot}{tr})$				

Dónde:

FV = Fuentes de Variación.

t = número de tratamientos.

GL= Grados de Libertad.

r = número de repeticiones.

SC = Suma de Cuadrados

y_i **=** Total del Tratamiento.

CM = Cuadrados Medios.

Y• = Gran Total

FCal = F Calculada

**** =** Rechazar H_0 usando $\alpha = 0.05$

FTab = F Tabular

*** =** Rechazar H_0 usando $\alpha = 0.01$

Trat = tratamientos

NS = No se rechaza Ho.

Sig = Significancia estadística

Cuadro 14. Tabla de recolección de datos, se anota el valor de Y de cada UE.

	R1	**R2**	**Rn**	**Totales**
T1	Y11	Y12	Y1n	Y1•
T2	Y21	Y22	Y2n	Y2•
Tn	Yn1	Yn2	Ynn	Y3•
				Y••

Se dice que el Análisis de la Varianza es Significativo cuando se rechaza la Hipótesis Nula que establece que los tratamientos producen el mismo efecto sobre la variable respuesta. Esto implica que la Fcal > Ftab o Pr < α.

Modelo Estadístico

$$Y_{ij} = \mu + T_i + E_{ij}$$

Dónde:
i = 1, 2,…, tratamientos.
j = 1, 2,…, repeticiones.
T_i = Efecto del tratamiento i.

y_{ij}=Es el valor de la variable respuesta del tratamiento i en su repetición j.
μ = Media General.
E_{ij} = Error experimental

Contraste de hipótesis.

Ho: Todos los tratamientos producen el mismo efecto sobre la variable respuesta.

Ha: Al menos dos tratamientos producen efectos diferentes sobre la variable respuesta.

Regla de decisión

Rechazar Ho Si Fcal > Ftab

Rechazar Ho Si P-Value < α

Ejemplo del Diseño Experimental Completamente Aleatorizado (DECA)

En la ciudad de Toluca se estableció un ensayo de procedencias de *Pinus radiata* y se evaluaron 3 lugares de procedencia (**T1** = Amecameca, Estado de México, **T2** = Ixmiquilpan, Hidalgo y **T3** = Cd. Madera, Chihuahua). Después de 20 años, se midió el diámetro de los árboles para observar si existía diferencia entre procedencias. Los datos se muestran en el Cuadro 15.

Cuadro 15. Datos obtenidos del experimento de ensayo de procedencias, tres tratamientos con 6 repeticiones.

	R_1	**R_2**	**R_3**	**R_4**	**R_5**	**R_6**
T1	43.3	42.9	40.5	42.5	33.5	43.4
T2	28.9	29.6	31.0	35.6	33.3	39.1
T3	38.0	48.5	37.3	37.5	33.5	33.4

Tratamientos
T_1 = Amecameca, Estado de México (Testigo).
T_2 = Ixmiquilpan, Hidalgo.
T_3 = Ciudad Madera, Chihuahua.

Variable Respuesta: Diámetro de los árboles a los 20 años (cm).

Análisis de la Varianza

El análisis de la varianza en el diseño experimental completamente aleatorizado nos permite identificar si entre los tratamientos existe diferencia de acuerdo con la variable de respuesta.

Contraste de hipótesis

Ho: Todas las procedencias producen el mismo efecto sobre el diámetro de *Pinus radiata.*
Ha: Al menos dos tratamientos producen efectos diferentes sobre el diámetro de *Pinus radiata*.

Para poder rechazar o no rechazar la Ho es necesario realizar el Análisis de la varianza. Para el ejemplo, los resultados se muestran en el Cuadro 16.

Cuadro 16. Desarrollo del cuadro del análisis de la varianza (DECA).

FV	GL	SC	CM	Fcal	Probabilidad	Ftab	Sig
Tratamiento	2	201.4	100.69	4.998	0.022	3.68	*
Error	15	302.2	20.15				
Total	17	503.6					

Suma del Total por Tratamientos
$Y_{1\cdot}$ = 246.1
$Y_{2\cdot}$ = 197.5
$Y_{3\cdot}$ = 228.2
Tratamientos = 3
Repeticiones = 6

Gran total
$Y_{\cdot\cdot}$ = 671.8

Grados de libertad

GL (Tratamientos) = T-1 = 3-1 = 2

GL (Total) = Tr-1 = 3(6)-1 = 18-1 = 17
GL (Error) = GL (Total)- GL (Tratamientos)=17-2=15

Suma de Cuadrados de Tratamientos

SC(Tra) = $\frac{y_{1\bullet}{}^2}{6} + \frac{y_{2\bullet}{}^2}{6} + \frac{y_{3\bullet}{}^2}{6} - \frac{y_{\bullet\bullet}{}^2}{Tr}$

$$=\frac{246.1^2}{6} + \frac{197.5^2}{6} + \frac{228.2^2}{6} - \frac{671.8^2}{18}$$

=10094.2016 + 6501.0416 + 8679.2066 - 25,073.0688
=201.381

Suma de Cuadrados del Total

SCtotal = $y^2_{11} + y^2_{12} + y^2_{13} + y^2_{14} + y^2_{15} + y^2_{16} + y^2_{21} + y^2_{22} + y^2_{23} + y^2_{24} + y^2_{25} + y^2_{26} + y^2_{31} + y^2_{32} + y^2_{33} + y^2_{34} + y^2_{35} + y^2_{36}.$

$$= 43.3^2 + 42.9^2 + 40.5^2 + 42.5^2 + 33.5^2 + 43.4^2 + 28.9^2 + 29.6^2 + 31.0^2 + 35.6^2 + 33.3^2 + 39.1^2 + 38.0^2 + 48.5^2 + 37.3^2 + 37.5^2 + 33.5^2 + 33.4^2 - \frac{671.8}{18^2}$$

SCtotal = 25 576.64 – 25 073.06889 = 503.571

Suma de Cuadrados del Error

SC(Error) = SC(total) - SC(Tra)

SC(Error) = 503.571-201.381

SC(Error) = 302.19

Cuadrados Medios

$$\boldsymbol{Cuadrado\ Medio\ de\ Tratamientos} = \frac{SC(TRA)}{GL(TRA)} = \frac{201.381}{2} = 100.69$$

$$\boldsymbol{Cuadrado\ Medio\ del\ Error} = \frac{SC(Error)}{GL(Error)} = \frac{302.190}{15} = 20.15$$

F calculada

$$\boldsymbol{Fcalculada} = \frac{CM(TRA)}{CM(ERROR)} = \frac{100.69}{20.15} = 4.998$$

F tabular

$$\boldsymbol{Ftab} = F_{GLError,\ \alpha}^{GL\ (Tra)}$$

$\boldsymbol{Ftab} = F_{15,\ 0.05}{}^{2} = 3.68$

Cálculo de F tabular utilizando Excel

1. Situarse en una celda.
2. Insertar la función $\boxed{\text{=Distr.F.inv. (0.05,2,15)}}$

= Distr.F.inv. (Probabilidad, Grado de libertad 1, Grado de libertad, 2)

Dónde:

Probabilidad = α (0.05, 0.01).

Grados de Libertad 1 = Grados de Libertad de Tratamientos.

Grados de Libertad 2 = Grados de Libertad del Error.

Cálculo de F tabular en tablas de Fisher.

	Prob valor alto F	GL (Trat) (2)
G (error) (15)	A 0.05	3.68

Conclusión

Para poder concluir el análisis de la varianza es necesario tomar en cuenta la siguiente:

Regla de Decisión.

"Rechazar Ho, si F calculada > F tabular"
"Rechazar Ho, si Probabilidad de Error < α "

Del Cuadro 16, se obtienen los valores tanto de F calculada, F tabular y Probabilidad.

F Calculada		**F tabular**
4.998	>	3.68

P- vale		**α**
0.022	<	0.05

De acuerdo con la regla se decisión, se rechaza Ho con α = 0.05. Por lo tanto con un nivel de significancia del 5 %(Confiabilidad del 95 %) se concluye que no todas las procedencias estudiadas producen el mismo efecto sobre el diámetro de los árboles de *Pinus radiata* después de 20 años.

Si en lugar de α = 0.05 tomamos α = 0.01. Entonces de acuerdo con la regla se decisión, no se rechaza Ho con α = 0.01. Por lo tanto con un nivel de significancia del 1 %(Confiabilidad del 99 %) se concluye que todas las procedencias estudiadas producen el mismo efecto sobre el diámetro de los árboles de *Pinus radiata* después de 20 años.

5.1 Coeficiente de correlación

El coeficiente de correlación es un parámetro que se utiliza para determinar si existe relación entre 2 variables de estudio conforme estas cambian sus valores. Por ejemplo, algunas veces es de interés conocer como es la relación del diámetro y el volumen de los árboles de distintas edades, por lo cual es conveniente calcular este parámetro.

El valor que toma el coeficiente de correlación puede ir de -1 hasta 1. Los valores muy cercanos a -1 y 1 indican una relación muy estrecha. Por ejemplo si el valor obtenido en el procedimiento de correlación es de -0.95, quiere decir que conforme se incrementa el valor de una variable, el valor de la otra variable disminuirá. Cuando el valor de la correlación es de 0.95 indica que cuando se incrementa el valor de un variable también se incrementa el valor de la otra.

Es aconsejable utilizar valores de correlación que sean mayores a 0.90 y menores a -0.90. Conforme los valores se acercan a cero, la relación entre las variables disminuye y no es aconsejable utilizar dicho valor de correlación.

La fórmula de para el cálculo del coeficiente de correlación es la siguiente

$$\boldsymbol{Coeficiente\ de\ Correlaci}ó\boldsymbol{n}\ (\boldsymbol{r}): \frac{\sum_{i=1}^{n}(Xi - Xmedia)(Yi - Ymedia)}{\sqrt{\sum_{i=1}^{n}(Xi - Xmedia)^2}\sqrt{\sum_{i=1}^{n}(Yi - Ymedia)^2}}$$

Donde:

r **=** Coeficiente de correlación

Xi, Yi **=** Valor de cada uno de los datos de la variable de interés

X media, Y media = Promedio de los valores de cada grupo de datos

Ejemplo

En el Parque Estatal de Monte Alto se realizó una evaluación de la situación actual de un predio que cuenta con arbolado de *Pinus pringlei*. Con la finalidad de caracterizar el comportamiento de los árboles, de acuerdo con el diámetro y la altura, se solicita el cálculo del Coeficiente de Correlación de Pearson.

Número de Árbol	Diámetro (cm) (X)	Altura (m)(Y)
1	43	21
2	38.5	20
3	7	1.8
4	5.5	1.54
5	5.5	1.5
6	5	1.15
7	5	1
8	4.3	0.77
9	3.4	0.6
10	1.9	0.26

Se recomienda realizar una tabla en donde se desarrolle el proceso de cálculo (Cuadro 17). Para el ejemplo tomaremos el Diámetro como X y la Altura como Y, pero es importante aclarar que puede tomarse cualquier variable como X o Y.

Cuadro 17. Desarrollo del procedimiento de cálculo del Coeficiente de Correlación.

Árbol	Diámetro (cm)	Altura (m)	X_i-Xmedia	Y_i-Ymedia	(X_i-Xmedia)*(Y_i-Ymedia)	(X_i-Xmedia)^2	(Y_i-Ymedia)^2
1	43	21	31.09	16.03	498.62	966.58	257.21
2	38.5	20	26.59	15.03	399.86	707.02	226.14
3	7	1.8	-4.91	-3.16	15.52	24.10	9.99
4	5.5	1.54	-6.41	-3.42	21.93	41.08	11.71
5	5.5	1.5	-6.41	-3.46	22.19	41.08	11.98
6	5	1.15	-6.91	-3.81	26.34	47.74	14.53
7	5	1	-6.91	-3.96	27.37	47.74	15.69
8	4.3	0.77	-7.61	-4.19	31.90	57.91	17.57
9	3.4	0.6	-8.51	-4.36	37.12	72.42	19.02
10	1.9	0.26	-10.01	-4.70	47.06	100.20	22.10
				Suma	1127.94	2105.929	605.99

Con los datos generados en la tabla podemos calcular el Coeficiente de Correlación (r).

$$\boldsymbol{Coeficiente\ de\ Correlaci}\text{ó}\boldsymbol{n}\ (\boldsymbol{r}): \frac{\sum_{i=1}^{n}(Xi - Xmedia)(Yi - Ymedia)}{\sqrt{\sum_{i=1}^{n}(Xi - Xmedia)^2}\ \sqrt{\sum_{i=1}^{n}(Yi - Ymedia)^2}}$$

$$\boldsymbol{Coeficiente\ de\ Correlaci}\text{ó}\boldsymbol{n}\ (\boldsymbol{r}): \frac{1127.94}{\sqrt{2105.92} * \sqrt{605.99}}$$

$$\boldsymbol{Coeficiente\ de\ Correlaci}\text{ó}\boldsymbol{n}\ (\boldsymbol{r}): \frac{1127.94}{45.89 * 24.62}$$

$$\boldsymbol{Coeficiente\ de\ Correlaci}\text{ó}\boldsymbol{n}\ (\boldsymbol{r}): \frac{1127.94}{1129.67} = 0.998$$

Interpretación

De acuerdo con el valor de correlación obtenido se interpreta lo siguiente:

La relación, entre las variables de diámetro y altura, es lineal positiva, es decir, en *Pinus pringlei* al incrementarse el diámetro también se incrementa la altura, siguiendo una tendencia lineal.

5.1.1 Cálculo de coeficiente de correlación mediante software estadístico

En un análisis de calidad de semillas de *Pinus greggii* se obtuvieron medidas de características físicas y se requiere verificar si existe relación entre esas variables (Cuadro 18).

Cuadro 18. Atributos físicos de la semilla de *Pinus greggii* cosechado en el huerto semillero del Rancho la Esperanza, Villa Victoria, Estado de México, Cosecha: 01-2011.

No de Semilla	Largo (mm)	Ancho (mm)	Grosor (mm)
Semilla 1	6.8	3.2	2
Semilla 2	6.8	3.4	2
Semilla 3	6.7	2.9	1.7
Semilla 5	5.7	2.9	1.6
Semilla 6	6.1	3.4	1.9
Semilla 7	6.3	2.9	1.8
Semilla 8	7.2	3.1	2

Para realizar el análisis de correlación en Excel® es necesario activar, en primer lugar, las funciones de herramientas en la pestaña de datos, para ello se siguen los siguientes pasos.

1. Abrir Excel, Clic en la pestaña Archivo y seleccionar la casilla de Opciones.
2. Dentro de la pestaña Opciones seleccionar la casilla Complementos, dentro de esta ventana, en la casilla Administrar, seleccionar Complementos de Excel y dar Clic en la Pestaña Ir.
3. En el Cuadro de Diálogo que se despliega, seleccione la Casilla Herramientas para Análisis y dar Clic en Aceptar.

Una vez realizado este procedimiento debe aparecer, en la pestaña de Datos, la Opción Análisis de Datos (Figura 18).

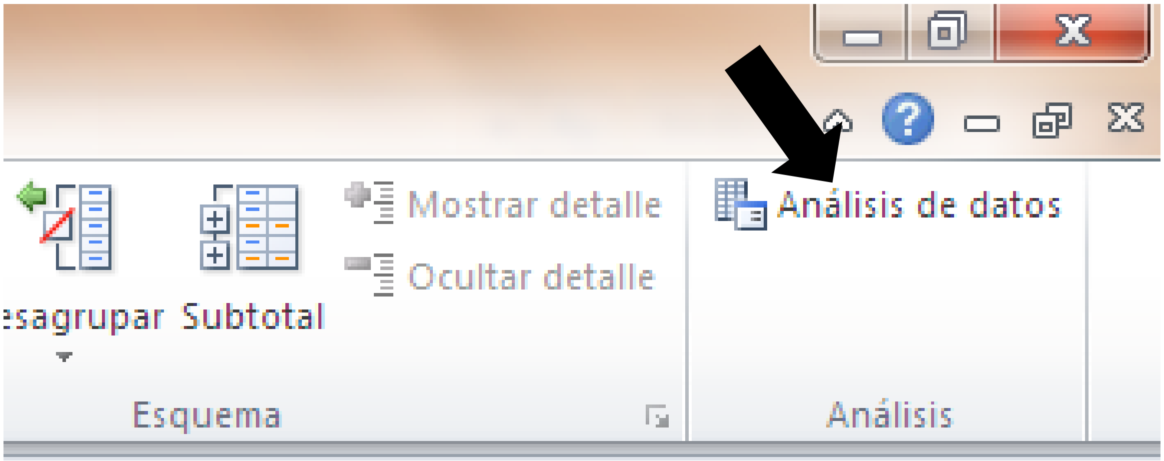


Figura 18. Apariencia de la pestaña Datos después de la activación de las herramientas de análisis.

Procedimiento de Correlación

1. Introducir los datos de las variables de interés en la hoja de Cálculo de Excel® y de la pestaña de Datos seleccionar Herramientas para análisis y dentro de ella Seleccionar Coeficiente de Correlación.

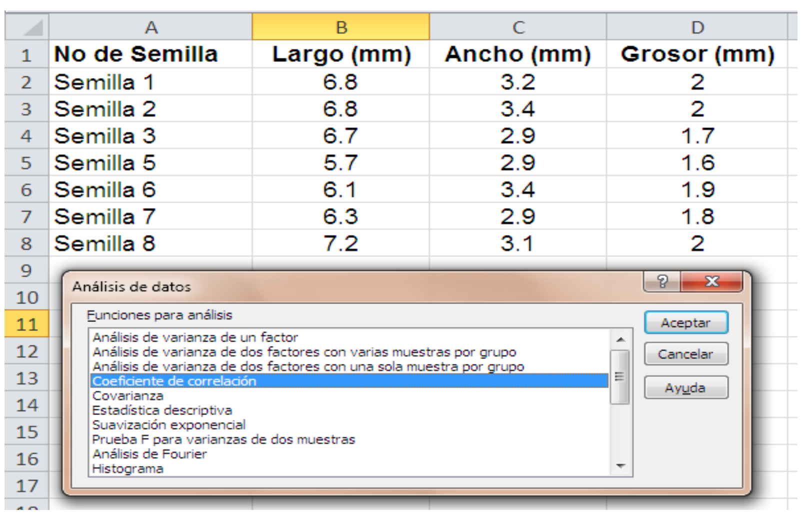
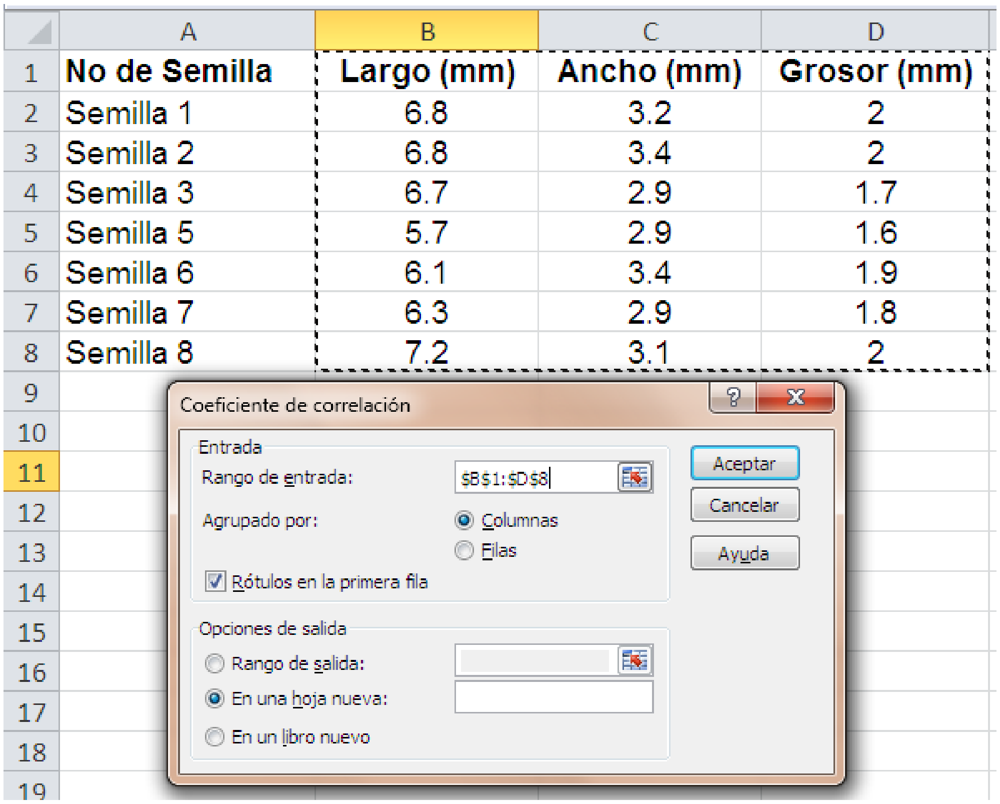


Figura 19. Introducción de datos y selección del Coeficiente de Correlación.

2. En el cuadro de diálogo que se despliega introducir los datos incluyendo los títulos como se muestra en la Figura 20. Es importante verificar que si las variables están colocadas en Columnas se selecciona en el cuadro de diálogo esta casilla y si están en Filas se selecciona esta casilla. En este caso están en Columnas.



No de Semilla	Largo (mm)	Ancho (mm)	Grosor (mm)
Semilla 1	6.8	3.2	2
Semilla 2	6.8	3.4	2
Semilla 3	6.7	2.9	1.7
Semilla 5	5.7	2.9	1.6
Semilla 6	6.1	3.4	1.9
Semilla 7	6.3	2.9	1.8
Semilla 8	7.2	3.1	2

Figura 20. Apariencias de la selección de datos para la correlación.

3. Al dar Click en la pestaña aceptar Excel® arrojará los resultados en una hoja nueva Figura 21.

	Largo (mm)	*Ancho (mm)*	*Grosor (mm)*
Largo (mm)	1		
Ancho (mm)	0.229393687	1	
Grosor (mm)	0.718192222	0.746061488	1

Figura 21. Resultado del análisis de Correlación de las variables de la semilla de *Pinus greggii.*

5.1.2 Interpretación del Análisis de Correlación

De acuerdo con los resultados se observa que estadísticamente no existe relación entre las variables Ancho y Largo de la Semilla (0.2229) ya que el coeficiente es muy bajo. La relación Grosor y Largo de la semilla (0.718) tampoco se considera buena, ya que, el coeficiente está por debajo de lo recomendado. Finalmente, la relación Grosor y Ancho (0.746) no es buena puesto que está por debajo del valor recomendado (0.90).

Ejemplo realizado con los datos de árboles de un predio en la Reserva Estatal de Monte Alto.

Se realizó el levantamiento de datos de Altura y Diámetro de los árboles de un predio de 500 m^2, ubicado en la reserva Estatal de Monte Alto, Valle de Bravo, Estado de México (Anexo 1), se realizó el procedimiento de correlación de se obtuvo el siguiente resultado (Cuadro 19).

Cuadro 19. Coeficiente de Correlación entre las variables Diámetro y Altura de los árboles de la Reserva Estatal de Monte Alto.

	Diámetro (cm)	**Altura (m)**
Diámetro (cm)	1	
Altura (m)	0.906113716	1

Interpretación: De acuerdo con los resultados obtenidos en el análisis de correlación, se observa que existe una relación directa positiva entre la altura y el diámetro de los árboles, de manera que al incrementarse la altura de los árboles también existe un incremento en diámetro. Esta información nos ayudará a generar un modelo matemático para poder estimar las variables sin necesidad de tomar la muestra en campo.

En la Figura 22 se puede observar el comportamiento de los datos y la relación que existe entre ellos. Nótese que al incrementarse los valores de la altura también se incrementa el diámetro.

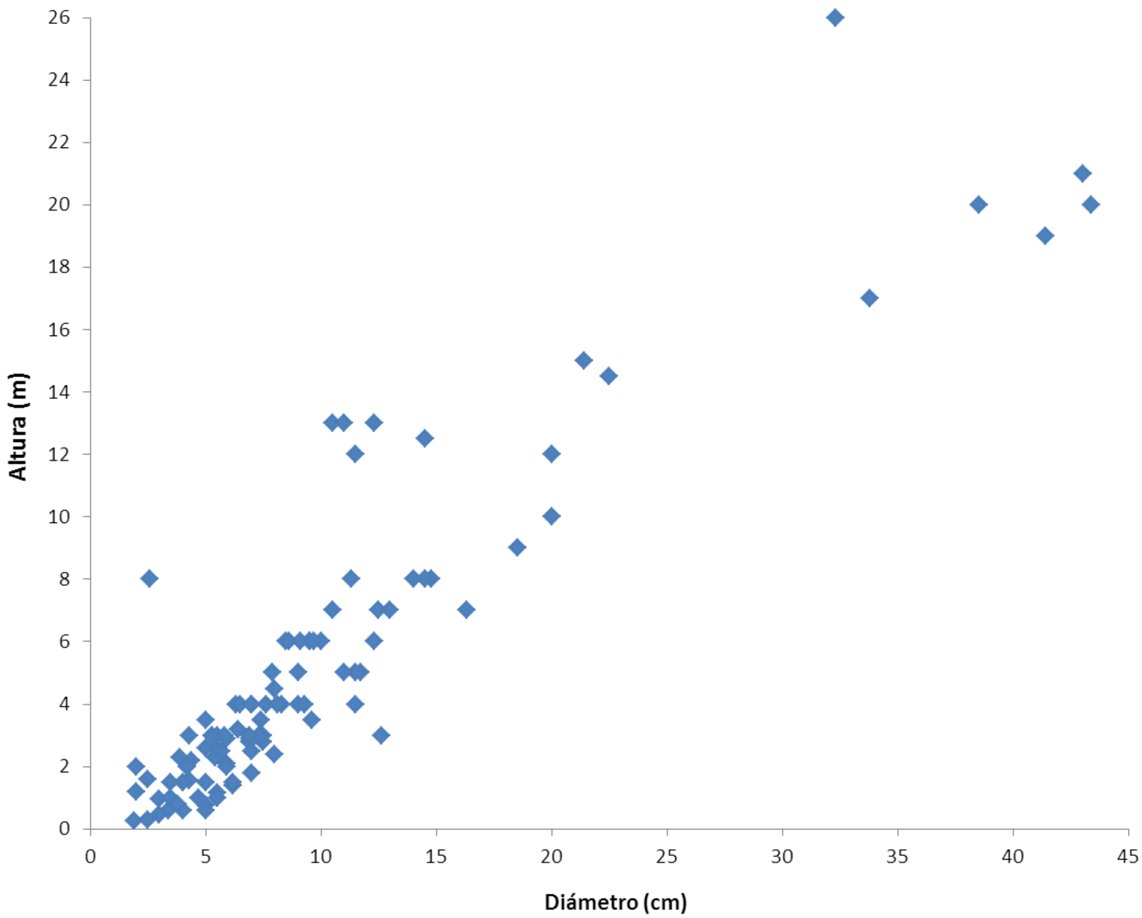


Figura 22. Grafica de dispersión de los datos de altura de diámetro de árboles de *Pinus pringleii.*

5.2 Covarianza

La Covarianza es una medida de asociación entre dos características que llamaremos X y Y.

Tiene las siguientes propiedades:

a) Cuando los valores de la variable X crecen con los de la variable Y, la Covarianza es Positiva.

b) Cuando los valores de la variable X decrecen al aumentar los de la variable Y, la covarianza es negativa.

Para el cálculo de la covarianza se utiliza la siguiente fórmula:

$$Sxy = \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i - x_{media}) - (y_i - y_{media})\right]$$

Dónde:

Sxy = Covarianza

n = Es el total de datos de X o Y

X = Variable de interés de estudio

Y = Variable de interés de estudio

Xmedia = Promedio de los valores de X

Y media = Promedio de los valores de Y

Para el cálculo de este parámetro es recomendable generar un cuadro en donde se desarrollarán los siguientes aspectos.

Ejemplo.

Utilizando los datos del Anexo 1 tendremos lo siguiente (Cuadro 20):

Cuadro 20. Desarrollo del cuadro para el cálculo de la Covarianza entre las variables de diámetro y altura de los árboles de *Pinus pringlei.*

Árbol	Diámetro (cm)	Altura (cm)	x-media	y-media	(x-media)(y-media)
1.0	43.0	21.0	33.1	15.6	517.8
2.0	38.5	20.0	28.6	14.6	418.8
3.0	5.5	1.2	-4.4	-4.2	18.5
4.0	5.0	1.5	-4.9	-3.9	18.9
5.0	5.0	0.8	-4.9	-4.6	22.5
6.0	3.4	0.6	-6.5	-4.8	31.0
97.0	10.0	6.0	0.1	0.6	0.1
98.0	8.6	6.0	-1.3	0.6	-0.9
99.0	9.1	6.0	-0.8	0.6	-0.5
100.0	9.7	6.0	-0.2	0.6	-0.1
Media	**9.9**	**5.4**		$\sum$	**3954.2**

Con el resultado de la suma aplicamos la ecuación y obtenemos la covarianza.

$$\boldsymbol{Sxy} = \frac{1}{n-1}\left[\sum_{i=1}^{n}(x_i - x_{media}) * (y_i - y_{media})\right]$$

$$\boldsymbol{Sxy} = \frac{1}{99}[3954.2]$$

$$\boldsymbol{Sxy} = 0.0101 * [3954.2]$$

$$\boldsymbol{Sxy} = 39.94\ \%$$

En este caso se concluye que si existe relación directa entre las variables de diámetro y altura de los árboles, ya que la Sxy es positiva y mayor que cero.

5.2.1 Interpretación de los resultados de la covarianza.

Cuando la Covarianza ($Sxy > 0$) hay dependencia directa (positiva), es decir, conforme se presentan grandes valores de X también se presentan o corresponden grandes valores de Y.

Cuando la Covarianza ($Sxy < 0$) hay dependencia inversa (negativa), es decir, conforme se presentan grandes valores de X se presentan o corresponden pequeños valores de Y.

Cuando la Covarianza ($Sxy = 0$) indica que no existe relación lineal entre las dos variables estudiadas.

VI. Intervalos de confianza

6.1 Conceptualización

Estimación: El objetivo principal de la estadística inferencial es la estimación, esto es que mediante el estudio de una muestra de una población se quiere generalizar las conclusiones hacia el total de dicha población.

Existen dos tipos de estimaciones para parámetros; puntuales y por intervalo.

La estimación puntual calcula un único valor estadístico que se usa para estimar un parámetro. El estadístico usado se denomina estimador.

Por ejemplo: Para conocer la altura de los 200 árboles de una plantación se toma una muestra de 10 árboles y se obtiene el promedio.

Promedio de los árboles = 2.5 m

Con este dato se infiere que la altura de todos los árboles es de 2.5 m aproximadamente.

La estimación por intervalo utiliza un rango, generalmente de ancho finito, que se espera que contenga el parámetro.

Por ejemplo: En una muestra de 30 semillas de *Pinus greggii* se analizaron variables físicas de calidad, como la longitud, ancho y grosor. Con estos datos se calculó el intervalo de confianza (Límite inferior y Límite superior) para estimar el comportamiento de la población de tal manera que el 95 % de las semillas quedaran dentro del intervalo de confianza (Cuadro 20).

Cuadro 20. Intervalo de confianza de una muestra de 30 semillas de *Pinus greggii.*

	Variable de interés (mm)		
Parámetro	**Longitud**	**Ancho**	**Grosor**
Limite Superior	6.71	3.18	1.97
Límite inferior	6.24	2.97	1.84

Tomando como ejemplo la longitud de la semilla donde el intervalo quedó como**:** **Límite superior:** 6.71 mm y **Límite Inferior:** 6.24 mm. Con esta información podemos estimar que, en una muestra de 100 semillas, la longitud del 95 de ellas caerá dentro de los valores de intervalo (entre 6.24 y 6.71). Esto porque se estimó con una probabilidad de error del 5 % y una confiabilidad del 95 %.

6.2 Cálculo del intervalo de confianza para la media

Un intervalo de confianza está constituido por 2 límites, el inferior y el superior. Como base se utiliza la media aritmética de los datos, la desviación estándar y el total de datos y se calcula con la siguiente fórmula.

$$\boldsymbol{Intervalo\ para\ la\ media} = \boldsymbol{x_{media}} \pm \boldsymbol{Z}\left(\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}\right)$$

Dónde:

Xmedia = Media aritmética de X.

σ = Desviación estándar de X.

n = Número total de datos.

Z = 1.96 (es el valor de Z cuando el intervalo se calcula al 95 % de confiabilidad) y 2.575 (es el valor de Z cuando el intervalo se calcula al 99 % de confiabilidad).

Cuadro 21. Parámetros calculados a partir de la longitud de semillas de *P. patula,* que se utilizarán para calcular los intervalos de confianza al 95 % y 99 %.

Número de Semilla	Longitud (mm)
1	5.7
2	5.2
3	5.9
4	5.5
5	5.2
6	5.5
7	5
8	5.2
9	5.2
10	4.9
Desviación Estándar	0.31
Promedio	5.33
Total de Datos	10

$$\boldsymbol{Intervalo\ para\ la\ media}\ (\mathbf{95}\%) = \boldsymbol{x_{media}} \pm \boldsymbol{Z}(\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}})$$

$$\boldsymbol{Intervalo\ para\ la\ media}\ (\mathbf{95}\%) = \mathbf{5.33} \pm \mathbf{1.96}(\frac{\mathbf{0.31}}{\sqrt{\mathbf{10}}})$$

$$\boldsymbol{Intervalo\ para\ la\ media}\ (\mathbf{95}\%) = \mathbf{5.33} \pm \mathbf{0.19214}$$

Límite superior: 5.33 + 0.19214 = 5.52214

Límite inferior: 5.33 – 0.19214 = 5.13786

$$\boldsymbol{Intervalo\ para\ la\ media}\ (\mathbf{99}\%) = \boldsymbol{x_{media}} \pm \boldsymbol{Z}(\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}})$$

$$\boldsymbol{Intervalo\ para\ la\ media}\ (\mathbf{99}\%) = \mathbf{5.33} \pm \mathbf{2.575}(\frac{\mathbf{0.31}}{\sqrt{\mathbf{10}}})$$

$$\boldsymbol{Intervalo\ para\ la\ media}\ (\mathbf{99}\%) = \mathbf{5.33} \pm \mathbf{0.2524}$$

Límite superior: 5.33 + 0.2524 = 5.5824

Límite inferior: 5.33 – 0.2524 = 5.0776

6.2.1 Interpretación del intervalo de confianza

De acuerdo con la probabilidad de error se concluye que:

Con una confiabilidad del 95 %, el valor de la longitud de todas las semillas de *P. patula* se encuentra entre 5.52214 y 5.13786.

Con una confiabilidad del 99 %, el valor de la longitud de todas las semillas de *P. patula* se encuentre entre 5.5824 y 5.0776.

Nota: Observe que, con una confiabilidad del 99 % el intervalo se hace más amplio, esto es porque al disminuir la probabilidad de error necesitamos asegurar que de cada 100 semillas, la longitud de 99 de ellas caiga dentro del intervalo.

Anexo 1. Variables de diámetro y altura de los árboles de *Pinus pringlei* de una predio de la reserva de Monte Alto.

Árbol	Diámetro (cm)	Altura (m)	Árbol	Diametro (cm)	Alt (m)
1	43.0	21.0	51	5.3	3.0
2	38.5	20.0	52	12.3	13.0
3	5.5	1.2	53	7.3	3.0
4	5.0	1.5	54	5.9	2.1
5	5.0	0.8	55	11.5	12.0
6	3.4	0.6	56	14.5	12.5
7	1.9	0.3	57	4.4	2.2
8	5.5	1.0	58	18.5	9.0
9	7.0	1.8	59	5.9	2.0
10	41.4	19.0	60	20.0	10.0
11	4.3	1.5	61	5.0	2.6
12	3.0	1.0	62	11.3	8.0
13	3.0	0.5	63	13.0	7.0
14	43.4	20.0	64	6.3	4.0
15	33.8	17.0	65	8.1	4.0
16	7.5	2.8	66	8.5	6.0
17	5.9	2.9	67	2.0	2.0
18	5.5	2.6	68	12.6	3.0
19	4.7	1.0	69	11.5	4.0
20	6.2	1.4	70	5.4	2.3
21	32.3	26.0	71	11.5	5.0
22	12.5	7.0	72	6.4	3.2
23	11.7	5.0	73	6.5	4.0
24	9.5	6.0	74	7.0	2.5
25	6.9	2.9	75	2.0	1.2
26	9.0	4.0	76	6.9	3.0
27	9.5	6.0	77	7.9	5.0
28	9.6	3.5	78	14.5	8.0
29	6.9	2.8	79	7.0	4.0
30	7.4	3.5	80	10.5	7.0
31	5.5	3.0	81	5.7	2.5
32	7.5	3.0	82	12.3	6.0
33	3.9	2.3	83	7.0	4.0
34	5.3	3.0	84	20.0	12.0
35	9.0	5.0	85	4.2	2.0
36	5.8	3.0	86	3.5	1.0
37	21.4	15.0	87	14.0	8.0

38	22.5	14.5	88	2.6	8.0
39	10.5	13.0	89	4.0	1.5
40	4.3	3.0	90	11.0	5.0
41	11.0	13.0	91	8.0	2.4
42	5.0	3.5	92	5.0	0.6
43	6.2	1.5	93	4.0	0.6
44	7.6	4.0	94	3.8	0.8
45	2.5	1.6	95	2.5	0.3
46	9.3	4.0	96	3.5	1.5
47	8.0	4.5	97	10.0	6.0
48	8.3	4.0	98	8.6	6.0
49	14.8	8.0	99	9.1	6.0
50	16.3	7.0	100	9.7	6.0