

# **Estadística Aplicada**

## **A la Forestería I**

**Ing. Felipe de Mendiburu**

---

### **ESTADÍSTICA**

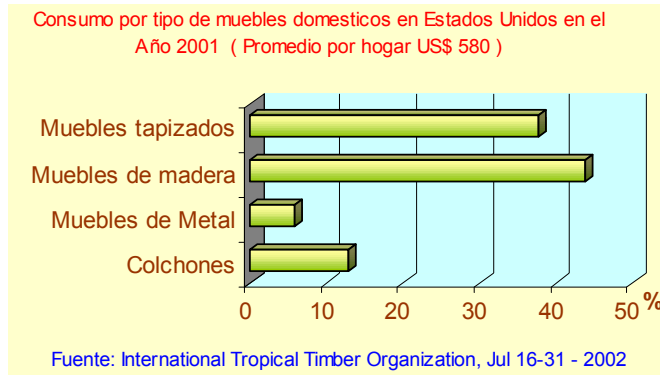
**Cuando se habla de estadística, se suele pensar:**

---

- **En una relación de datos numéricos presentada en un texto escrito, en forma ordenada y sistemática.**
- **En un estudio de fenómenos mediante la descripción del mismo o a través de inferencias con distribuciones probabilísticas .**
- **En ciencia con métodos y procedimientos que permiten realizar la caracterización, análisis e interpretación de una serie de datos para la toma de decisiones frente a situaciones de incertidumbre**

## Estadística

Cuando se presentan gráficos de datos:



## Estadística

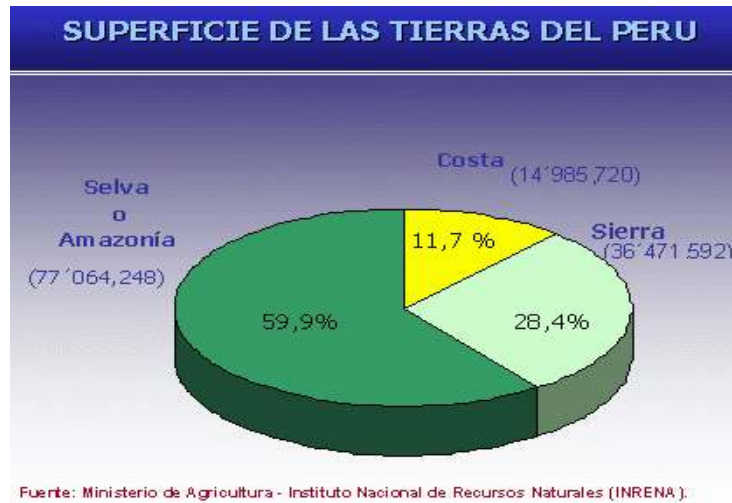
Cuando se presentan tablas de datos:

<b>PRODUCCIÓN DE MADERA</b>		
<b>En miles de metros cúbicos.</b>		
Tipo	1998	1999
Madera aserrada	590.30	834.67
Parqué	5.87	3.66
Madera contrachapada	134.00	34.32
Láminas de madera	0.71	3.66
Chapas decorativas	0.40	0.00
Durmientes	2.14	2.24
Carbón	28.19	136.29
Postes	5.31	1.01
Total	766.90	1015.90

Fuente: Instituto Nacional de Estadística e Informática.

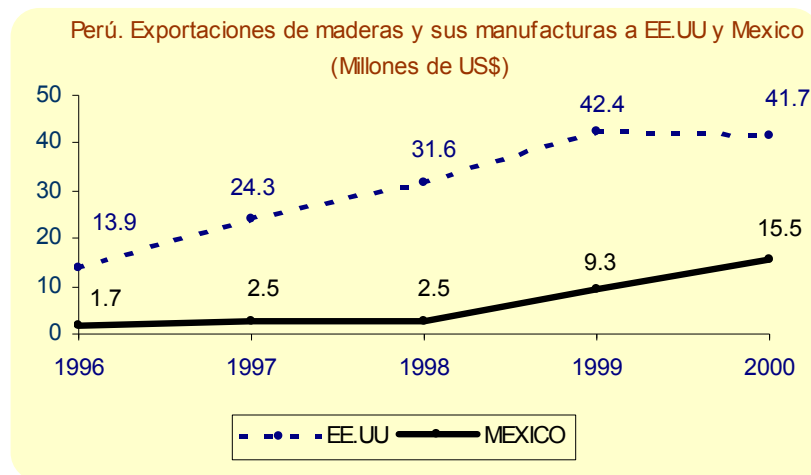
## Estadística

Cuando se presentan resúmenes de datos:



## Estadística

Cuando se presentan proyecciones comparativas de datos:



## Estadística

### Cuando se presentan textos de datos:

---

- “ El Año 1997 fue declarado Año de la reforestacion en el Perú, el area plantada ese año fue de 109 890 has. Los años anteriores fueron:

1996	57 448 Has.
1995	28 345 "
1994	36 221 "
1993	27 999 "
1992	7 804 "

Fuente INRENA-1997

- “..el costo del transporte de materia prima (madera aserrada) es muy elevado, representando entre el 25 y el 30% del precio de la misma puesta en su región de origen.”
- “..Las plantaciones existentes en la sierra en 1992 comprendían 278.498 hectáreas distribuidas entre los departamentos de Cusco (21,3%), Junín (12,3%), Apurímac (11,9%), Cajamarca (11,4%) y Ancash (10%). Estas plantaciones han sido realizadas en un 80% en base a *Eucalyptus Globulus* ...“

## Conceptos de estadística

---

- ✓ **Individuos o elementos:** Personas, empresas, especies de arboles, parcelas agrícolas u objetos que contienen cierta información que se desea estudiar.
- ✓ **Población:** conjunto de individuos o elementos que cumplen ciertas propiedades comunes. El estudio de toda la población constituye un CENSO.
- ✓ **Muestra:** subconjunto representativo de una población.
- ✓ **Parámetro:** función definida sobre los valores numéricos de características medibles de una población.
- ✓ **Estadístico:** función definida sobre los valores numéricos de una muestra.

### Tamaño de la población, ésta puede ser:

---

- ✓ **Finita**, es contables, pequeña. Por ejemplo el número de productores de madera aserrada.
- ✓ **Infinita**, No es contable, grande. Por ejemplo el numero de compradores de muebles en madera.

### Estudio de la población:

---

- ✓ **Observación**, Es el registro que se obtiene al evaluar una característica en una unidad elemental.
- ✓ **Variable**, Es toda característica de las unidades elementales sujetas a evaluación, mediante medición, conteo o calificación .

### Tipos de Variables :

---

**Variables Cuantitativas**.- Son aquellas que generan observaciones de carácter numérico y son del tipo:

- **Discretas**.- Cuando el conjunto de todas las posibles observaciones que se generan constituyen a lo más un conjunto infinito numerable.
- **Continuas**.- Cuando el conjunto de todas las posibles observaciones que se generan constituye un conjunto infinito no numerable.

#### Ejemplos:

- Número de postes producidas por día (V.C.D.)
- Ancho de las vigas. (V.C.C.)
- Tamaño de parcela del agricultor. (V.C.C)
- Exportación de madera, miles de dolares. (V.C.C)

Continuación...

## Tipos de Variables :

---

**Variables Cualitativas.**- Son aquellas que generan observaciones no numéricas y son del tipo:

- **Jerárquicas - cuasi cuantitativas.**- Cuando se puede establecer una relación de orden entre las posibles observaciones .
- **Nominales.**- No existe un orden entre las posibles observaciones.

Ejemplos:

- Calidad de las maderas laminadas producidas. (V.C.J.)
- Color de la madera. (V.C.N.)
- Especies maderables. (V.C.N.)
- Producción de madera por año. (V.C.J.)

## Medidas estadísticas :

---

**Parámetro.**- Es una constante que describe una característica de una población. Para su medición se requiere todos los datos de la población.

- **De tendencia central o de resumen:**
  - La media o promedio. ( $\mu$ )
  - La mediana. (Me)
  - La moda. (Mo)
  - La media geométrica. (Mg).
- **De dispersión o variabilidad:**
  - La variancia o varianza. ( $\sigma^2$ )
  - La desviación estándar. ( $\sigma$ )
  - El coeficiente de variabilidad. (C.V.)

## Medidas estadísticas :

---

**Valor Estadístico**.- es un valor análogo al parámetro, pero es calculado con la información de la muestra.

Los valores estadísticos son variables, pueden tomar diferentes valores al cambiar de muestra.

Un valor estadístico estima a un parámetro.  $s^2$  estima a  $\sigma^2$

- Promedio  $\bar{X}$
- Mediana (me)
- Moda (mo)
- Variancia  $s^2$
- Desviación estándar s

## Etapas para un estudio estadístico:

---

### 1. Formular o definir el problema

Delimitar con exactitud que debe ser estudiado, población sobre lo que se va a concluir.

### 2. Diseño o Plan experimental.

Definir si se va a realizar un censo o un muestreo.

### 3. Compilar los datos.

Establecer los procedimientos e instrumentos a usar.

### 4. Organizar y describir los datos (Ordenar, Clasificar y Calcular)

### 5. Decisión e Inferencia final.

Cuando se ha realizado un muestreo y es necesario concluir sobre la población, se debe realizar la estimación de parámetros y la comprobación de hipótesis sobre los mismos.

## Casos para estudio:

1. La industria de la madera está compuesta básicamente por la producción de madera aserrada, cerca del 90 por ciento del total, en orden de importancia sigue la fabricación de triplay, parquet, chapas, entre otras.
2. La madera aserrada continua como el principal producto exportado, destacando la caoba, su participación en el total exportado se ha ido incrementándose hasta representar el 67,9 por ciento en el año 2000.

**Poblacion:** Productores en la industrialización de la madera.

**Muestra:** 20 productores elegidas para el estudio.

**Unidad:** Un productor

**Variable:** Producción de madera aserrada mensual, **Tipo:** cuantitativa continua.  
**Observación:** 90 m3

**Parámetro:** promedio de producción por productor en madera aserrada.

**Valor del parámetro:** desconocido. **Valor estadístico:** promedio de producción por productor en madera aserrada de los 20 productores, **Valor:** 100 m3

## Organización de datos :

### Tablas estadísticas

La población o muestra de  $n$  individuos, es agrupada en 'K' clases, según carácter  $c_1, c_2, \dots, c_k$ . Para cada una de las clases se determina:

**Frecuencia absoluta.**- ( $f_i$ ) Numero de individuos en la clase 'i'

**Frecuencia relativa.**-( $fr_i$ )

**Frecuencia absoluta acumulada** ( $F_i$ )

**Frecuencia relativa acumulada** ( $Fr_i$ )

$$fr_i = \frac{f_i}{n} \quad \sum_{i=1}^k fr_i = fr_1 + fr_2 + \dots + fr_k = 1 \quad F_i = \sum_{k=1}^i f_k \quad Fr_i = \frac{F_i}{n} = \sum_{k=1}^i fr_k$$

Clase	frecuencia	fr	F	Fr
c1	f 1	fr 1	F 1	Fr 1
c2	f 2	fr 2	F 2	Fr 2
ck	f k	fr k	F k	Fr k

## Elección de las clases:

---

En cuanto a la elección de las clases, debe seguir los siguientes criterios en función del tipo de variable:

- **Variables cualitativas o cuasicuantitativas**, las clases serán de tipo nominal según los valores de la variable.
- **Variables cuantitativas**, existen dos posibilidades:
  - Si la variable es discreta, las clases serán valores numéricos  $x_1, x_2, \dots, x_k$
  - Si la variable es continua las clases serán definidas por **intervalos**. Las clases están determinadas por los intervalos. En la tabla se agrega una columna para escribir la marca de clase.

## Elección de intervalos para variables continuas:

---

Se fija el número de intervalos

$$N^{\circ} \text{ intervalos} \equiv k \approx \begin{cases} \sqrt{n} & \text{si } n \text{ no es muy grande,} \\ 1 + 3,22 \log n & \text{en otro caso.} \end{cases}$$

Se determina el **Rango o Amplitud** = Máximo – Mínimo valor

Luego el **Tamaño de Intervalo de clase (TIC)** = Rango / k

Finalmente los **límites y la marca de clase**.

$Li(1)$  = Mínimo valor; los siguientes se agrega el TIC

$Ls(1) = Li(2) = Li(1) + TIC$

Marca de clase =  $(Li(i) + Ls(i)) / 2$

Cada intervalo es cerrado por la izquierda y abierto por la derecha.

### **Ejemplo para variable continua:**

Se muestreo 100 arboles de bolaina en su fase de adaptacion. A las 12 meses se midio su altura, diametro en centimetros. Para el presente ejemplo se trabajo con la altura, teniendo los siguientes resultados:

$$\text{Minimo} = 60$$

$$\text{Maximo} = 330$$

$$\text{Amplitud} = 330 - 60 = 270$$

$$\text{Nro. de intervalos} = k = 1 + 3.22 \text{ Log}(100) = 7$$

$$\text{TIC} = 270/7 = 38,6$$

$$\text{Li1} = 60; \quad \text{Ls1} = 60 + 38,6 = 98,6$$

$$\text{Li2} = 98,6; \quad \text{Ls2} = 137,2$$

$$\text{Marca de clase 1} = c1 = (60+98,6)/2 = 79,3$$

### **Tabla de Frecuencia: Altura en (cm) de arboles de capirona en la primera evaluación a las 12 meses :**

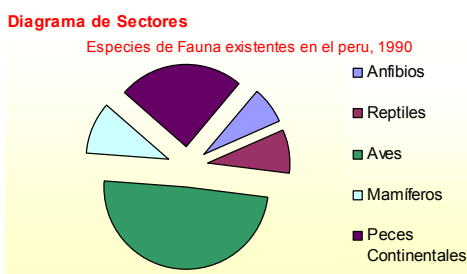
Intevalo		M.Clase	fi	fri	Fi	Fri
60	98.6	79.3	1	0.01	1	0.01
98.6	137.2	117.9	7	0.07	8	0.08
137.2	175.8	156.5	9	0.09	17	0.17
175.8	214.4	195.1	16	0.16	33	0.33
214.4	253	233.7	30	0.3	63	0.63
253	291.6	272.3	25	0.25	88	0.88
291.6	330.2	310.9	12	0.12	100	1
			100	1		

## Gráficos para variables de estudio:

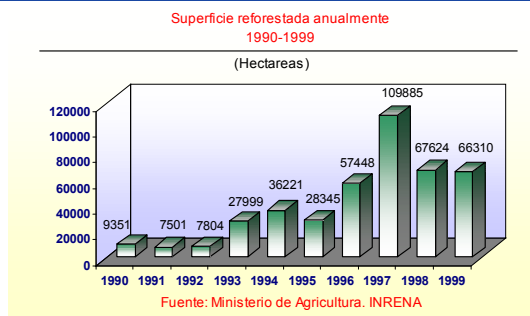
Principales diagramas según el tipo de variable	
Tipo de variable	Diagrama
V. Cualitativas	Barra, Sectores, Pictogramas
V. Discretas	Diferencial (barras)
	Integral (en escalera)
V. Continua	Diferencial (Histograma, Poligono de frecuencia)
	Integral (diagramas acumulados)

## Ejemplo para variable Cualitativa:

**Nominal:**

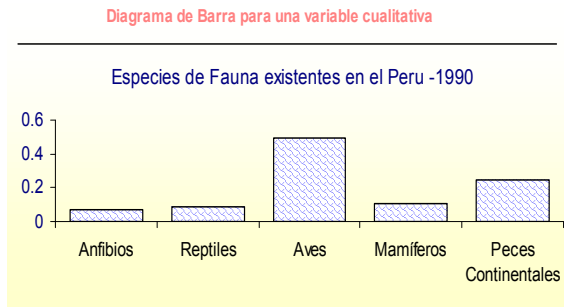


**Jerarquico:**

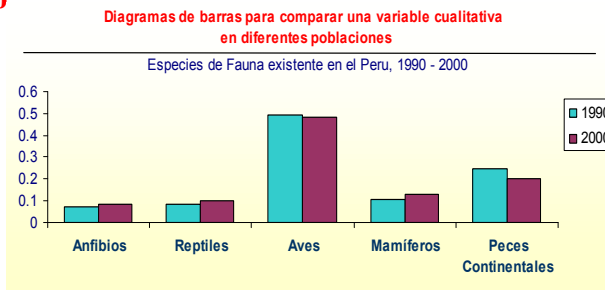


### Ejemplo para variable Cualitativa:

#### Barras:

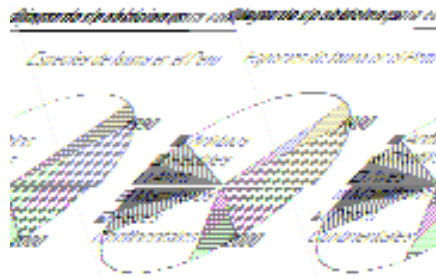


#### Comparativo:



### Ejemplo para variable Cualitativa:

#### Comparativo:



#### Pictograma:



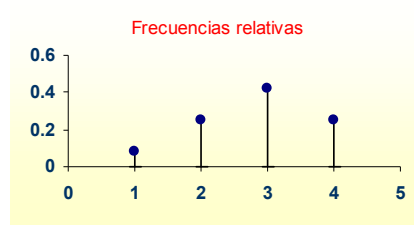
**Ejemplo para variable Cuantitativa Discreta:**

Variable: Número de hijos por familia, muestra de 12 familias

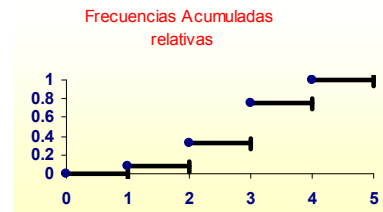
**Histograma:**

Hijos	$f_i$	$fr_i$	$F_i$	$Fr_i$
1	1	0.08	1	0.08
2	3	0.25	4	0.33
3	5	0.42	9	0.75
4	3	0.25	12	1
	12	1		

**Digrama Diferencial:**

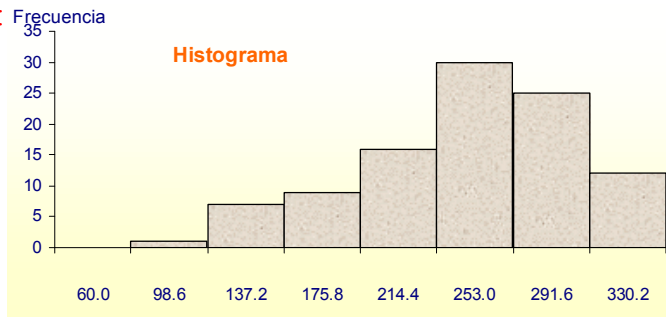


**Diagrama Integral:**

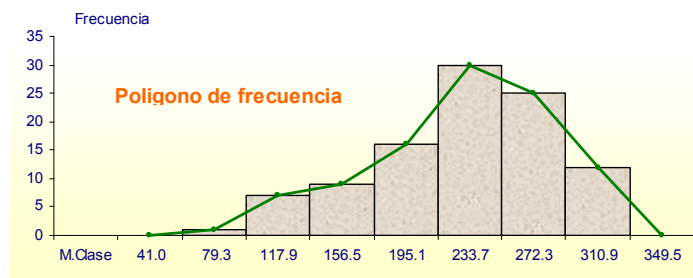


**Ejemplo para variable Cuantitativa Continua: Diferencial**

**Histograma:**

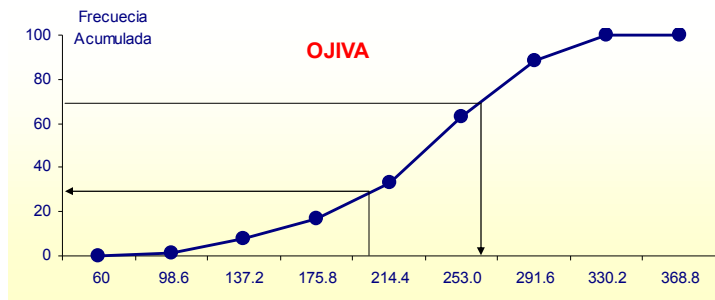


**Poligono:**



## Ejemplo para variable Cuantitativa Continua: Integral

### Ojiva:



### Usos:

1. Para cálculo de frecuencias en intervalos
2. Para cálculo de probabilidades
3. Para modelos de Simulación

## Medidas de tendencia central :

Media o Promedio Aritmético: Representa al conjunto de datos de la población o muestra

$$\text{Promedio}_{\text{aritmético}} = \frac{\sum_{i=1}^n x_i}{n} = \text{AVERAGE}(C4:C23)$$

### Continuos

- Promedio anual de producción de parquet
- Promedio mensual de exportación en madera laminar en m3

### Discretos

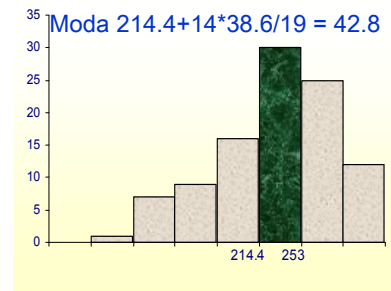
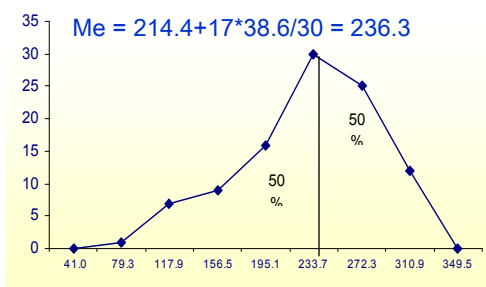
- Promedio de número de especies de árboles maderables por productor
- Promedio de número de especies de árboles maderables por productor.
- Promedio mensual del número de agricultores que solicita crédito en el presente año

## Medidas de tendencia central :

**Mediana.** Es el punto medio que separa la población o muestra en dos grupos con igual número de observaciones. `MEDIAN(C4:C23)`

Se utiliza como valor representativo del conjunto cuando se tiene valores extremos.

**Moda.** Representa el valor más frecuente del conjunto de datos, es útil para datos discretos y continuos cuando se tiene valores extremos y es notorio la presencia de una sola moda. `MODE(C4:C23)`



## Medidas de tendencia central :

**Percentiles.** - Es el punto ( $P_p$ ) que divide al conjunto de datos en  $p\%$  menor o igual que el valor percentil y  $(1-p)\%$  mayores que el valor percentil.

$$P_p = L_i + \frac{np - F_{i-1}}{f_i} TIC \quad 0 < p < 1$$

donde:  $i$  = Clase percentil.

La clase percentil es el intervalo de clase donde se supera por primera vez los  $(np)$  datos,  $F_i > np$ .

$L_i$  = Límite inferior de la clase percentil

$F_{i-1}$  = Frecuencia acumulada absoluta de la clase anterior a la clase percentil.

$f_i$  = Frecuencia absoluta de la clase percentil

$P_{25} = Q_1$  (Primer cuartil)

$P_{75} = Q_3$  (tercer cuartil)

$P_{50} = Q_2 = Me$

En Excel,

`=PERCENTILE(C4:C23, 0.25)`

### Cálculo del P0.25 = Q1 primer cuartil:

Intevalo		M.Clase	fi	fri	Fi	Fri
60	98.6	79.3	1	0.01	1	0.01
98.6	137.2	117.9	7	0.07	8	0.08
137.2	175.8	156.5	9	0.09	17	0.17
175.8	214.4	195.1	16	0.16	33	0.33
214.4	253	233.7	30	0.3	63	0.63
253	291.6	272.3	25	0.25	88	0.88
291.6	330.2	310.9	12	0.12	100	1

- Se fija la clase del percentil 0.25
- Se aplica la formula del percentil.

$$P_{0.25} = 175.8 + \frac{100(0.25) - 17}{16} 38.6 = 195.1$$

El 25% de la muestra de bolaina tienen una altura menos de 195.1 cm. En este caso coincide con la marca de clase, por 25 esta en el medio entre 17 y 33.

### Medidas de tendencia central :

**Media geométrica.** - Es útil para promediar tasas, por ejemplo tasa promedio de producción anual de madera aserrada especie cedro. Años 1990-1999 .

$$Media \_ Geometrica = \sqrt[n]{x_1 x_2 \dots x_n}$$

#### Ejemplo:

Años	Cedro	Tasa
1990	44.7	
1991	35	0.78
1992	43.4	1.24
1993	51.3	1.18
1994	57.9	1.13
1995	53.4	0.92
1996	61.4	1.15
1997	35.7	0.58
1998	49	1.37
1999	61.5	1.25

Promedio =  $(0,78 \times 1,24 \times \dots \times 1,25)^{(1/9)}$  (^ = potencia)

En Excel: =GEOMEAN(C3:C11)

Tasa Promedio = 1.04

Se concluye que hay un crecimiento promedio del 4% anual en la producción de amdera aserrada desde 1990 a 1999.

## Medidas dispersion :

---

**Rango o amplitud:** Representa la máxima variabilidad de los datos de la población o muestra. Es una medida absoluta y tiene las mismas unidades de la variable.

$$Rango = Max(x) - Min(x)$$

### Continuos

- Amplitud anual de producción de parquet
- Amplitud mensual de exportación en madera laminar en m<sup>3</sup>

### Discretos

- Rango del número de especies de árboles maderables por productor
- Rango mensual del número de agricultores que solicita crédito en el presente año

## Medidas dispersion :

---

**Desviación Media:** Representa la variación promedio respecto a la mediana o el promedio. Se utiliza cuando se tienen valores extremos y la mediana es el valor más representativo. Es una medida absoluta y tiene las mismas unidades de la variable.

$$DM = \sum \frac{|x_i - Me|}{n}$$

### Continuos

- Desviación media anual de producción de madera laminar en lquitos, dado en m<sup>3</sup>

## Medidas dispersion :

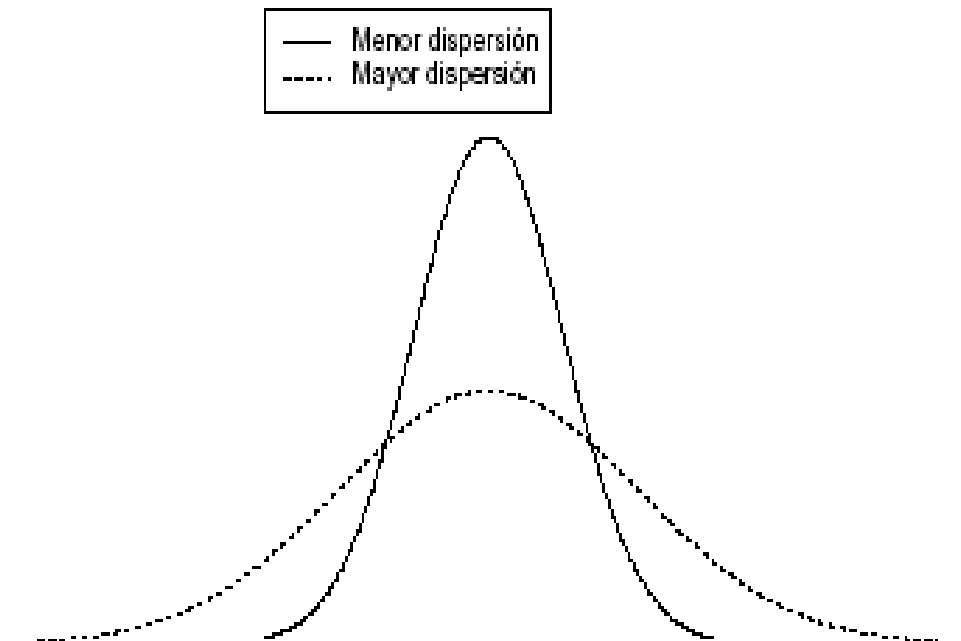
**Desviacion estandar:** Representa la variabilidad absoluta de los datos de la poblacion o muestra. Tiene las mismas unidades de la variable. Se utiliza cuando la poblacion o muestra no tiene valores extremos.

$$s = \sqrt{\sum \frac{(x_i - \text{media})^2}{n-1}}$$
$$\sigma = \sqrt{\sum \frac{(x_i - \text{media})^2}{N}}$$

### Continuos

- Desviacion estandar de produccion de parquet mensual en el año 2003

En poblacion normal, El 67% de los datos esta alrededor de la media en +/- una desviacion estandar



## Medidas dispersion :

---

**Variancia:** Medida de variacion cuadratica de la variable, el uso esta limitado para los analisis estadistico. Tambien es una medida de variacion absoluta. Su unidad es la unidad de la vaiable al cuadrado.

$$S^2 = \sum \frac{(x_i - \text{media})^2}{n-1} \quad \text{Poblacion}$$
$$\sigma^2 = \sum \frac{(x_i - \text{media})^2}{N} \quad \text{Muestra}$$

### Continuos

- Variancia de produccion de parquet

## Medidas dispersion :

---

**Coefficiente de Variacion CV:** Medida de variacion relativa, no tiene unidades y su uso es comparativo; por lo general se expresa en porcentaje. El mayor uso es para determinar tamaños de muestra, evaluar los experimentos de campo y de condiciones controladas..

$$CV = \frac{S * 100\%}{\text{promedio}}$$

### Continuos

- Coeficiente de variacion de experimentos sobre estudios de fertilizacion en arboles frutales.
- Medios controlados < 20 %
- Campo < 30 %

