CAPITULO 5

MEDIDAS DE DISPERSIÓN

No solo basta con determinar las medidas de tendencia central para comprender el comportamiento de una serie de datos, es importante además, conocer que tan alejados están esos datos respecto a ese punto de concentración.

Las medidas de dispersión nos indican la distancia promedio de los datos respecto a las medidas de tendencia central. Así podremos diferenciar dos conjuntos de datos que poseen iguales medias, siendo los datos de uno más dispersos del otro.

CAPITULO 5: MEDIDAS DE DISPERSIÓN

Medidas de dispersión: Son indicadores estadísticos que muestran la distancia promedio que existe entre los datos y la media aritmética.

En el estudio de las medidas de dispersión daremos un vistazo a cuatro indicadores básicos:

- Desviación media
- Varianza
- Desviación estándar
- Coeficiente de variación

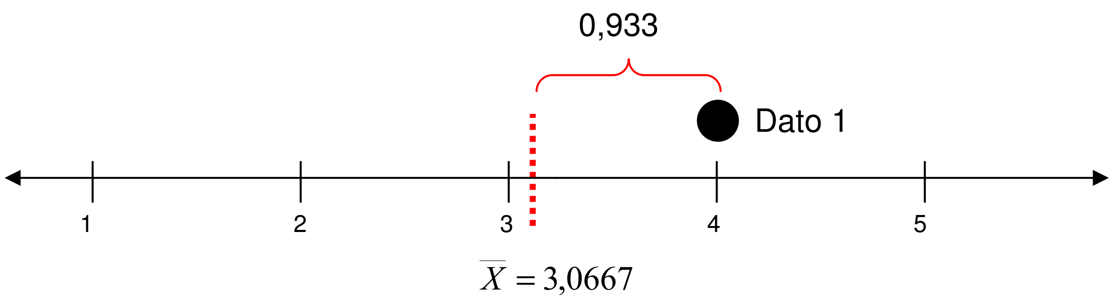
El cálculo de cada uno de ellos se toma basado en la media aritmética.

5.1 DESVIACIÓN MEDIA

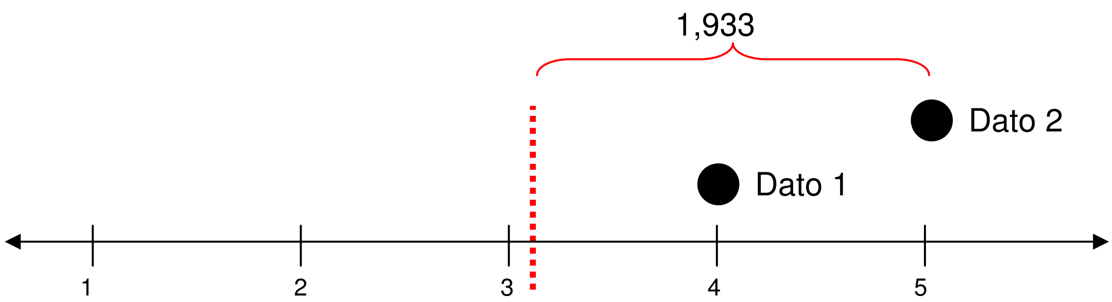
Para conocer con un solo indicador que tan disperso se encuentran un conjunto de datos a un punto de concentración, debemos como primera medida, calcular la distancia de cada dato respecto a una medida de tendencia central. Por ejemplo:

4	5	3
5	3	2
2	2	2
3	5	1
4	1	4

Tenemos que la media aritmética es de aproximadamente 3,0667 (indicador de tendencia central por excelencia). El primer dato (4), se aleja de la media en 0,9333 hacia la derecha. Gráficamente tendríamos:

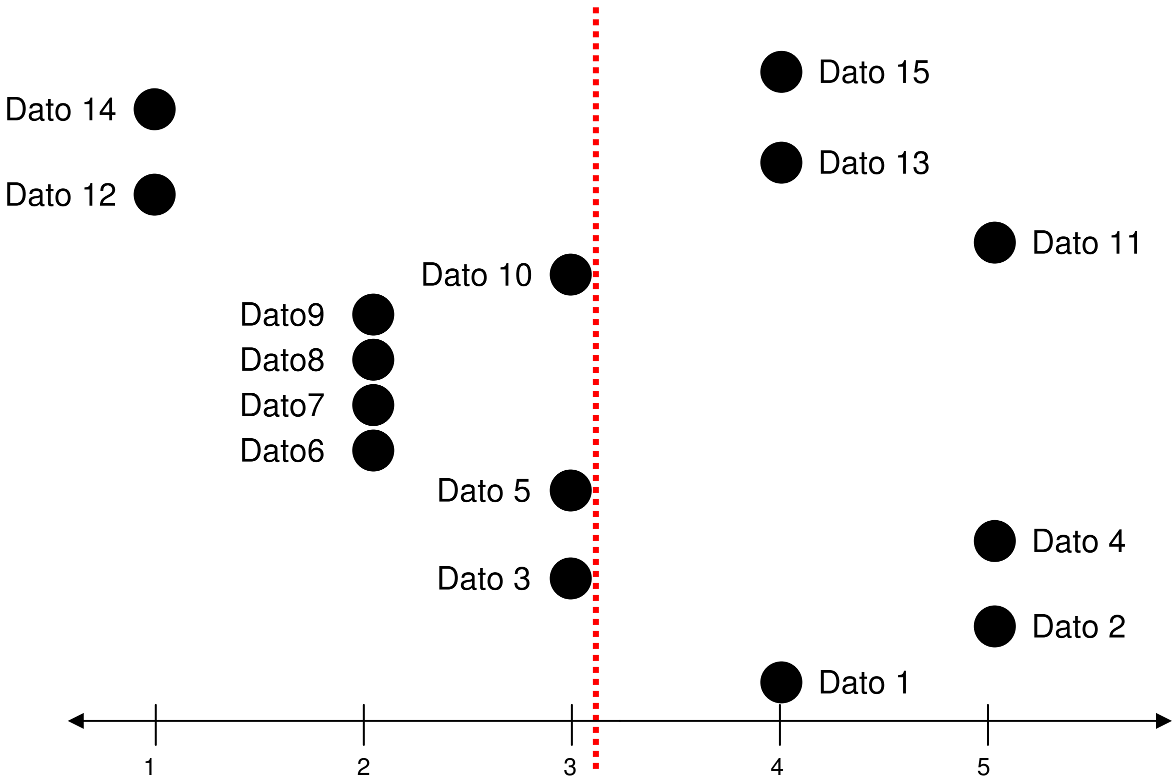


Para el segundo dato (5) la distancia es de 1,9333 respecto a la media aritmética:



$\overline{X} = 3{,}0667$

Note que el tercer dato (3) posee una distancia de 0,0667 hacia la izquierda de la media. Para indicar las distancias de estos puntos, agregaremos el signo negativo, por tanto, la distancia del tercer dato sería –0,0667. La representación gráfica de todos los puntos quedaría:



$\overline{X} = 3{,}0667$

El total de las distancias de los puntos que están a la izquierda respecto a la media es de -8,6 (empleando todos los decimales), que es igual a la sumatoria de las distancias de los puntos que están a la derecha respecto a la media 8,6. Concluimos que la sumatoria de todas las distancias de cada punto respecto a la media aritmética es igual a cero (las distancias se anulan):

$$\sum_{i=1}^{n}\left(X_i - \overline{X}\right) = 0$$

Para responder a la pregunta de ¿qué tan disperso están los datos respecto a la media aritmética?, recurriremos nuevamente al promedio simple. Para llegar a una fórmula básica de dispersión, en que las distancias positivas y negativas no se eliminen, modificaremos la fórmula anterior para trabajar solo con distancias positivas mediante el valor absoluto:

$$\sum_{i=1}^{n}\left|X_i - \overline{X}\right| = 17{,}2$$

La distancia promedio sería de aproximadamente 1,15 (resultado de la división entre la distancia total absoluta y el total de datos). A esta distancia promedio se le conoce con el nombre de desviación media y significa que en promedio, los datos se separan de la media en 1,15.

Desviación media (Dm): Equivale a la división de la sumatoria del valor absoluto de las distancias existentes entre cada dato y su media aritmética y el número total de datos.

$$Dm = \frac{\sum_{i=1}^{n}\left|X_i - \overline{X}\right|}{n}$$

Se debe hacer la distinción que para datos poblacionales (no agrupados), la fórmula quedaría:

$$Dm = \frac{\sum_{i=1}^{N}\left|X_i - \mu\right|}{N}$$

La variación para los datos agrupados en tablas tipo B radica en cambiar el valor de X_i por la marca de clase correspondiente, multiplicando esa distancia por su frecuencia:

$$Dm = \frac{\sum_{i=1}^{Nc} |Mc_i - \mu| . f_i}{N}$$

Población

$$Dm = \frac{\sum_{i=1}^{Nc} |Mc_i - \overline{X}| . f_i}{n}$$

Muestra

Para las tablas tipo A solo cambiaremos la marca de clase por su respectivo valor de clase (representada por X_i):

$$Dm = \frac{\sum_{i=1}^{Nc} |X_i - \mu| . f_i}{N}$$

Población

$$Dm = \frac{\sum_{i=1}^{Nc} |X_i - \overline{X}| . f_i}{n}$$

Muestra

5.1.1 Ejemplo: Desviación media para datos no agrupados

Tres alumnos son sometidos a una competencia para probar sus conocimientos en 10 materias diferentes, cada una sustentada con 10 preguntas. La idea del concurso es encontrar al alumno más idóneo para representar al colegio en un torneo a nivel nacional.

El número de preguntas buenas por materia se muestra a continuación:

Materia	Carlos	Pedro	Juan
1	2	7	5
2	9	2	6
3	10	2	5
4	2	6	5
5	3	6	5
6	1	3	5
7	9	6	4
8	9	7	5
9	1	6	6
10	4	5	4

SOLUCIÓN

Lo primero que analizaremos es la media de los puntajes para cada uno de los alumnos, con el fin de determinar el alumno con mayor promedio de preguntas buenas.

$$Carlos : \overline{X}_c = \frac{50}{10} = 5$$

$$Pedro : \overline{X}_p = \frac{50}{10} = 5$$

$$Juan : \overline{X}_j = \frac{50}{10} = 5$$

Las medias para los resultados de los alumnos coinciden: los tres alumnos tienen responden en promedio 5 preguntas correctas por prueba. ¿Cuál sería entonces el indicador diferenciador entre los alumnos?.

Complementemos el análisis anterior calculando la desviación media:

$$Dm_c = \frac{|2-5| + |9-5| + |10-5| + |2-5| + |3-5| + |1-5| + |9-5| + |9-5| + |1-5| + |4-5|}{10} = \frac{39}{10} = 3{,}9$$

$$Dm_p = \frac{|7-5| + |2-5| + |2-5| + |6-5| + |6-5| + |3-5| + |6-5| + |7-5| + |6-5| + |5-5|}{10} = \frac{21}{10} = 2{,}1$$

$$Dm_j = \frac{|5-5| + |6-5| + |5-5| + |5-5| + |5-5| + |5-5| + |4-5| + |5-5| + |6-5| + |4-5|}{10} = \frac{9}{10} = 0{,}9$$

Carlos muestra una desviación media de 3,9 indicando que los datos se alejan en promedio de la media en 3,9 preguntas buenas. Pedro disminuye su variación (2,9), siendo Juan el que menos variación presenta con 0,9 preguntas tanto por arriba como por debajo de la media aritmética. Se recomienda al colegio elegir como ganador en este caso a Juan, presenta resultados más constantes que los otros dos alumnos, Juan en promedio acierta 5 preguntas buenas con una variación muy baja (rondando entre 4 y 6).

5.1.2 Ejemplo: Desviación media para datos agrupados

Una maquina dispensadora de gaseosas esta programada para llenar un envase con 350 c.c. de un refresco popular. A partir de una muestra de prueba realizada sobre 30 envases se realizó la siguiente tabla de frecuencia:

Ni	Lm	Ls	F	Mc
1	130.0	140.1	2	135.1
2	140.1	150.1	5	145.1
3	150.1	160.1	14	155.1
4	160.1	170.1	4	165.1
5	170.1	180.1	4	175.1
6	180.1	190.0	1	185.1
Total			**30**	

Calcular e interpretar la desviación media.

SOLUCIÓN

PASO 1: Calcular la media aritmética.

$$\overline{X} = \frac{4712{,}84}{30} = 157{,}095$$

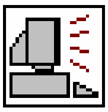
PASO 2: Calcular la desviación media.

$$Dm = \frac{|135{,}1-150{,}095|.2+|145{,}1-150{,}095|.5+|155{,}1-150{,}095|.14+|165{,}1-150{,}095|.4+|175{,}1-150{,}095|.4+|185{,}1-150{,}095|.1}{30}$$

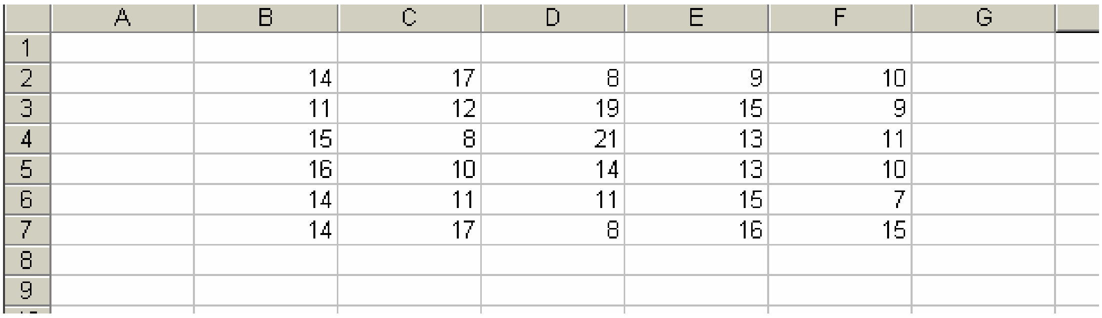
La desviación media es de aproximadamente 8,8 c.c. Concluimos que con datos suministrados de una muestra, el dispensador llenó los 30 envases con un promedio de 157,095 c.c. con una desviación media de 8,8 c.c.

La desviación media describe un rango de dispersión promedio de llenado del dispensador, ubicándolo entre 148,295 c.c. (equivale a restar la media a la desviación media) y 165,895 c.c. (sumar una desviación media a la media aritmética).

5.1.3 Cálculos de la desviación media en Excel



Presentaremos el cálculo de la desviación media en Excel tanto para datos sin agrupar, como para los datos agrupados en tablas de frecuencias. Copiemos los siguientes datos a partir de la celda B2.



	A	B	C	D	E	F	G
1							
2		14	17	8	9	10	
3		11	12	19	15	9	
4		15	8	21	13	11	
5		16	10	14	13	10	
6		14	11	11	15	7	
7		14	17	8	16	15	
8							
9							

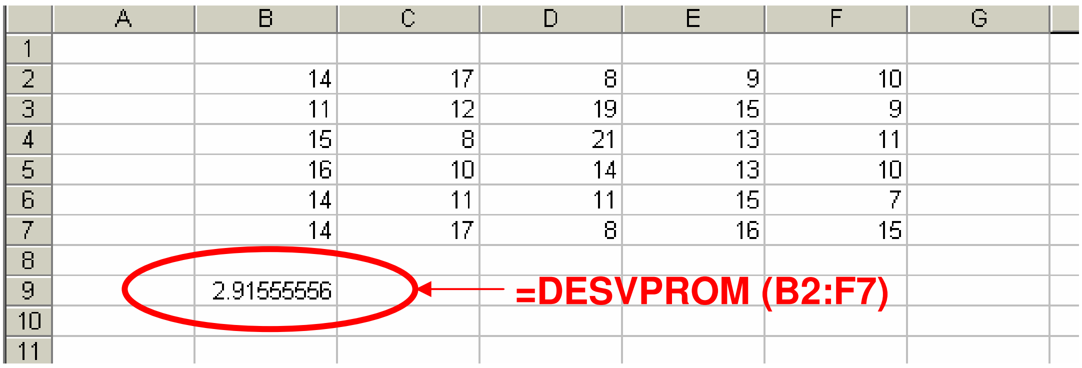
Excel cuenta con la función **DESVPROM** para el cálculo de la desviación media para datos sin agrupar.

DESVPROM: Calcula la desviación media de un conjunto de datos numéricos.

Formato: DESVPROM(número1;número2;...)
Categoría: Estadísticas

Activemos esta función en la celda B9, señalando el rango de celdas B2:F7 en el campo número1.

Al pulsar en el botón Aceptar, se mostrará la desviación media.



	A	B	C	D	E	F	G
1							
2		14	17	8	9	10	
3		11	12	19	15	9	
4		15	8	21	13	11	
5		16	10	14	13	10	
6		14	11	11	15	7	
7		14	17	8	16	15	
8							
9		2.91555556					
10							
11							

Para el cálculo de la desviación media en tablas de frecuencia debemos calcular de antemano la media aritmética y el valor absoluto de las distancias.

Copiemos la siguiente tabla de frecuencia en una hoja nueva en Excel (es la misma utilizada en el ejemplo 5.1.2).

	A	B	C	D	E	F	G
1							
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	
3		1	130	140.1	2	135.1	
4		2	140.1	150.1	5	145.1	
5		3	150.1	160.1	14	155.1	
6		4	160.1	170.1	4	165.1	
7		5	170.1	180.1	4	175.1	
8		6	180.1	190	1	185.1	
9		**Total**			**30**		
10							
11							

El primer paso es calcular la media aritmética para datos agrupados con ayuda de la función **SUMAPRODUCTO** (ver el ejemplo dado en el punto 4.1.7), aplicado sobre las frecuencias y marcas de clases.

	A	B	C	D	E	F	G
1							
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	
3		1	130	140.1	2	135.1	
4		2	140.1	150.1	5	145.1	
5		3	150.1	160.1	14	155.1	
6		4	160.1	170.1	4	165.1	
7		5	170.1	180.1	4	175.1	
8		6	180.1	190	1	185.1	
9		**Total**			**30**		
10							
11		157.1					
12							

=SUMAPRODUCTO(E3:E8,F3:F8)/E9

Luego hallaremos las distancias de cada marca de clase respecto a la media, convirtiéndolas a su valor absoluto con la función **ABS**.

> **ABS:** Devuelve el valor absoluto de un número.
>
> **Formato:** ABS (número)
> **Categoría:** Matemáticas y trigonométricas

Esta función posee un único campo (número) el cual contendrá, la distancia entre la marca de clase y la media. Para el primer intervalo de clase tendríamos:

=ABS(F3-B11)

Donde F3 representa la primera marca de clase y B11 la media aritmética. Para completar el cálculo, multiplicaremos esta función por la frecuencia respectiva:

=ABS(F3-B11)*E3

Para poder arrastrar la fórmula, debemos recordar que la celda B11 no varía (la media aritmética es una sola), ubicándonos sobre las letras B11 en modo de edición y luego pulsando la tecla F4.

=ABS(F3-B11)*E3

El resultado final, después de haber arrastrado la fórmula, debería verse como sigue:

	A	B	C	D	E	F	G	H
1								
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	**Distancia**	
3		1	130	140.1	2	135.1	44	
4		2	140.1	150.1	5	145.1	60	
5		3	150.1	160.1	14	155.1	28	
6		4	160.1	170.1	4	165.1	32	
7		5	170.1	180.1	4	175.1	72	
8		6	180.1	190	1	185.1	28	
9			**Total**		**30**		**264**	
10								
11		157.1						

El total de las distancias se muestra en la celda G9. La desviación (que ubicaremos en la celda B12), es el resulta de la división de la distancia total sobre el número de datos empleados en el ejercicio.

	A	B	C	D	E	F	G	H
1								
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	**Distancia**	
3		1	130	140.1	2	135.1	44	
4		2	140.1	150.1	5	145.1	60	
5		3	150.1	160.1	14	155.1	28	
6		4	160.1	170.1	4	165.1	32	
7		5	170.1	180.1	4	175.1	72	
8		6	180.1	190	1	185.1	28	
9			**Total**		**30**		**264**	
10								
11		157.1						
12		8.8						

=G9/E9

5.2 LA VARIANZA

Otra forma para asegurar que las diferencias entre la media y los puntos de un valor positivo, es elevándola al cuadrado. Al promedio de estas distancias al cuadrado se le conoce como varianza.

Varianza (S^2 o σ^2): Es el resultado de la división de la sumatoria de las distancias existentes entre cada dato y su media aritmética elevadas al cuadrado, y el número total de datos.

$$S^2 = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{n-1} \quad \text{o} \quad \sigma^2 = \frac{\sum_{i=1}^{N}\left(X_i - \mu\right)^2}{N}$$

Distinguimos dos símbolos para identificar la varianza: S^2 para datos muestrales, y σ^2 para datos poblacionales. Note que la fórmula para la varianza muestral presenta en su denominador al tamaño de la muestra menos uno, tendencia adoptada por los estadísticos para denotar una varianza más conservadora.

Al igual que ocurre con la desviación media, podemos definir las fórmulas para datos agrupados en tablas tipo A y tipo B. Para las tablas tipo A tenemos:

$$\sigma^2 = \frac{\sum_{i=1}^{Nc}\left(X_i - \mu\right)^2 . f_i}{N}$$

Población

$$S^2 = \frac{\sum_{i=1}^{Nc}\left(X_i - \overline{X}\right)^2 . f_i}{n}$$

Muestra

Para las tablas tipo B, la clase cambia por la marca de clase del intervalo:

$$\sigma^2 = \frac{\sum_{i=1}^{Nc}\left(Mc_i - \mu\right)^2 . f_i}{N}$$

Población

$$S^2 = \frac{\sum_{i=1}^{Nc}\left(Mc_i - \overline{X}\right)^2 . f_i}{n}$$

Muestra

Una advertencia en el uso de esta medida, es que al elevar las distancias al cuadrado, automáticamente se elevan las unidades. Por ejemplo, si unidad trabajada en los datos es centímetros, la varianza da como resultados centímetros al cuadrado.

5.2.1 Ejemplo: Varianza para datos no agrupados

La siguiente muestra representa las edades de 25 personan sometidas a un análisis de preferencias para un estudio de mercado.

25	19	21	35	44
20	27	32	38	33
18	30	19	29	33
26	24	28	39	31
31	18	17	30	27

Determinar la varianza.

SOLUCIÓN

PASO 1: Calcular la media aritmética.

$$\overline{X} = \frac{694}{25} = 27,76$$

PASO 2: Calcular la varianza

En este punto, la varianza es identificada por S^2.

$$S^2 = \frac{(25-27,76)^2 + (19-27,76)^2 + (21+27,76)^2 + ... + (27-27,76)^2}{25-1}$$

$$S^2 = \frac{1244,56}{24} = 51,8567$$

La varianza equivale a 51,8567. Por elevar las unidades al cuadrado, carece de un significado contextual dentro del análisis descriptivo del caso.

5.2.2 Ejemplo: Varianza para datos agrupados

Calcular la varianza a partir de la siguiente tabla de frecuencia (suponga que los datos son poblacionales).

Ni	Lm	Ls	f	Mc
1	[15	17)	2	16
2	[17	19)	5	18
3	[19	21)	13	20
4	[21	23)	4	22
5	[23	25]	1	24
Total			**25**	

SOLUCIÓN

PASO 1: Calcular la media aritmética.

$$\mu = \frac{16x2 + 18x5 + 20x13 + 22x4 + 24x1}{25} = \frac{494}{25}$$

$$\mu = 19{,}76$$

PASO 2: Calcular la varianza

En este punto, la varianza es identificada por S^2.

$$\sigma^2 = \frac{(16-19{,}76)^2.2 + (18-19{,}76)^2.5 + (20+19{,}76)^2.13 + (22-19{,}76)^2.4 + (24-19{,}76)^2.1}{25}$$

$$\sigma^2 = \frac{82{,}56}{25} = 3{,}3024$$

5.2.3 Cálculo de la varianza en Excel

Excel posee dos funciones propias para el cálculo de la media, diferenciando los datos muestrales de los datos poblacionales.

VAR: Calcula la varianza de una muestra.

Formato: VAR(número1;número2;…)
Categoría: Estadísticas

VARP: Calcula la varianza de todos los datos de una población.

Formato: VARP(número1;número2;…)
Categoría: Estadísticas

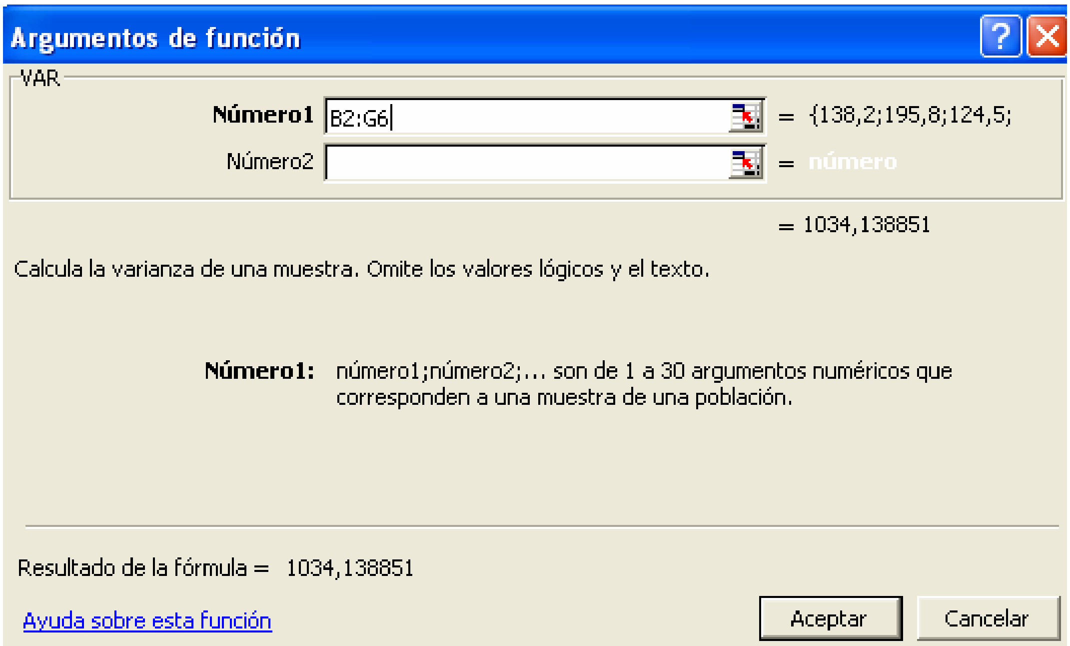
Mostremos su funcionamiento calculando la varianza en ambos casos a partir de los siguientes datos:

138,2	195,8	124,5	101,7	137,1	130,3
110,0	101,4	104,5	128,5	135,5	197,5
159,6	140,7	103,2	134,3	191	180,6
189,9	186,3	116,4	155,3	146,6	199,1
188,4	113,8	121,9	135,7	142,6	125,6

Los datos copiados en Excel desde la celda B2 deberían verse como sigue:

	A	B	C	D	E	F	G	H
1								
2		138,2	195,8	124,5	101,7	137,1	130,3	
3		110,0	101,4	104,5	128,5	135,5	197,5	
4		159,6	140,7	103,2	134,3	191	180,6	
5		189,9	186,3	116,4	155,3	146,6	199,1	
6		188,4	113,8	121,9	135,7	142,6	125,6	
7								
8								

Si los datos provienen de una muestra, emplearemos la función **VAR**, en cuyo denominador se tendría el valor 29 en vez de 30, equivalente al tamaño de la muestra. Activemos esta función en la celda B8.



El resultado de la varianza muestral es de 1034,138051.

	A	B	C	D	E	F	G	H
1								
2		138,2	195,8	124,5	101,7	137,1	130,3	
3		110,0	101,4	104,5	128,5	135,5	197,5	
4		159,6	140,7	103,2	134,3	191	180,6	
5		189,9	186,3	116,4	155,3	146,6	199,1	
6		188,4	113,8	121,9	135,7	142,6	125,6	
7								
8		1034,139		=VAR(B2:G6)				
9								

En la celda B9 calculemos la varianza para datos poblacionales.

	A	B	C	D	E	F	G	H
1								
2		138,2	195,8	124,5	101,7	137,1	130,3	
3		110,0	101,4	104,5	128,5	135,5	197,5	
4		159,6	140,7	103,2	134,3	191	180,6	
5		189,9	186,3	116,4	155,3	146,6	199,1	
6		188,4	113,8	121,9	135,7	142,6	125,6	
7								
8		1034,139						
9		999,6676						
10								

=VARP(B2:G6)

La función de la varianza **VARP**, divide la sumatoria de las distancias al cuadrado por los 30 datos, dando como resultado un valor menor que con la función **VAR** (la varianza para la muestra es un valor más conservador).

Para el cálculo de la varianza en datos agrupados en Excel, tomaremos la tabla de frecuencia dada en el ejemplo 5.2.2.

	A	B	C	D	E	F	G
1							
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	
3		1	15	17	2	16	
4		2	17	19	5	18	
5		3	19	21	13	20	
6		4	21	23	4	22	
7		5	23	25	1	24	
8		**Total**			**25**		
9							
10							

Calculemos la media en la celda B10.

	A	B	C	D	E	F	G
1							
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	
3		1	15	17	2	16	
4		2	17	19	5	18	
5		3	19	21	13	20	
6		4	21	23	4	22	
7		5	23	25	1	24	
8		**Total**			**25**		
9							
10		19,76					
11							

=SUMAPRODUCTO(E3:E7;F3:F7)/E8

En una columna adicional colocaremos las diferencias entre la marca de clase y la media elevadas al cuadrado multiplicadas por su frecuencia.

	A	B	C	D	E	F	G	H
1								
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	**Distancias**	
3		1	15	17	2	16	28,2752	
4		2	17	19	5	18	15,4880	
5		3	19	21	13	20	0,7488	
6		4	21	23	4	22	20,0704	
7		5	23	25	1	24	17,9776	
8			**Total**		**25**		**82,5600**	
9								
10		19,76						
11								

Analicemos la fórmula empleada desde la celda C3.

=(F3-B10)^2*E3

La celda B10 esta fija indicando la media aritmética. Aparece el operador ^, la cual eleva al cuadrado lo que esta dentro del paréntesis. Esta distancia se multiplica por el número de veces que se repite (por su frecuencia). Al final calculamos su sumatoria.

En la celda B11 calculamos la varianza.

	A	B	C	D	E	F	G	H
1								
2		**Ni**	**Lm**	**Ls**	**f**	**Mc**	**Distancias**	
3		1	15	17	2	16	28,2752	
4		2	17	19	5	18	15,4880	
5		3	19	21	13	20	0,7488	
6		4	21	23	4	22	20,0704	
7		5	23	25	1	24	17,9776	
8			**Total**		**25**		**82,5600**	
9								
10		19,76						
11		3,3024						

=G8/E8

5.3 DESVIACIÓN ESTÁNDAR

Habíamos visto que la varianza transforma todas las distancias a valores positivos elevándolas al cuadrado, con el inconveniente de elevar consigo las unidades de los datos originales.

La desviación estándar soluciona el problema obteniendo la raíz cuadrada de la varianza, consiguiendo así, un valor similar a la desviación media.

Desviación estándar o típica (S o σ): Es igual a la raíz cuadrada de la varianza.

$$S = \sqrt{S^2} \text{ o } \sigma = \sqrt{\sigma^2}$$

La S representa la desviación estándar de una muestra, mientras que σ la desviación para todos los datos de una población. Ampliando las fórmulas tenemos:

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

Población

$$S = \sqrt{\frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n}}$$

Muestra

Aplicamos el mismo procedimiento a las fórmulas para las tablas de frecuencias tipo A.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{Nc} (X_i - \mu)^2 . f_i}{N}}$$

Población

$$S = \sqrt{\frac{\sum_{i=1}^{Nc} \left(X_i - \overline{X}\right)^2 . f_i}{n}}$$

Muestra

Y para las tablas de frecuencias tipo B.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{Nc} (Mc_i - \mu)^2 . f_i}{N}}$$

Población

$$S = \sqrt{\frac{\sum_{i=1}^{Nc} \left(Mc_i - \overline{X}\right)^2 . f_i}{n}}$$

Muestra

5.3.1 Ejemplo: Desviación estándar para datos no agrupados

Calcular la desviación estándar al siguiente conjunto de datos muestrales.

220	215	218	210	210
219	208	207	213	225
213	204	225	211	221
218	200	205	220	215
217	209	207	211	218

PASO 1: Calcular la media aritmética.

$$\overline{X} = \frac{220 + 215 + 218 + 210 + 210 + ... + 218}{25} = \frac{5339}{25}$$

$$\overline{X} = 213{,}56$$

PASO 2: Calcular la varianza

En este punto, la varianza es identificada por S^2.

$$S^2 = \frac{(220 - 213{,}56)^2 + (215 - 213{,}56)^2 + (218 + 213{,}56)^2 + (210 - 213{,}56)^2 + ... + (218 - 213{,}56)^2}{25 - 1}$$

$$S^2 = \frac{1030{,}16}{25 - 1} = 42{,}9233$$

PASO 3: Calcular la desviación estándar a partir de la raíz cuadrada de la varianza.

$$S = \sqrt{42{,}9233}$$

$$S = 6{,}5516$$

Los datos se alejan en promedio de la media aritmética en 6,5516 puntos.

5.3.2 Ejemplo: Desviación estándar para datos agrupados

Calcular la desviación estándar a partir de la siguiente tabla de frecuencia. Considere los datos como poblacionales.

No.	Lm	Ls	f	Mc
1	13,20	15,21	15	14,21
2	15,21	17,21	10	16,21
3	17,21	19,21	1	18,21
4	19,21	21,21	4	20,21
5	21,21	23,21	5	22,21
6	23,21	25,21	12	24,21
7	25,21	27,20	1	26,21
Total			48	

PASO 1: Calcular la media aritmética.

$$\mu = \frac{14{,}21x15 + 146{,}21x10 + 18{,}21x1 + 20{,}21x4 + 22{,}21x5 + 24{,}21x12 + 26{,}21x1}{48} = \frac{902}{48}$$

$$\mu = 18{,}7917$$

PASO 2: Calcular la varianza

En este punto, la varianza es identificada por σ^2.

$$\sigma^2 = \frac{(14{,}21 - 18{,}7917)^2.15 + (16{,}21 - 18{,}7917)^2.10 + (18{,}21 - 18{,}7917)^2.1 + ... + (26{,}21 - 18{,}7917)^2.1}{48}$$

$$\sigma^2 = \frac{2789{,}96}{48} = 58{,}1242$$

PASO 3: Calcular la desviación estándar a partir de la raíz cuadrada de la varianza.

$$\sigma = \sqrt{58{,}1242}$$

$$\sigma = 7{,}6239$$

Los datos se alejan en promedio de la media aritmética en 7,6239 puntos.

5.3.3 Cálculo de la Desviación estándar en Excel

Al igual que en la varianza, Excel posee dos funciones para el cálculo de la media, diferenciando los datos muestrales de los datos poblacionales.

DESVEST: Calcula la desviación estándar de una muestra.

Formato: DESVEST(número1;número2;...)
Categoría: Estadísticas

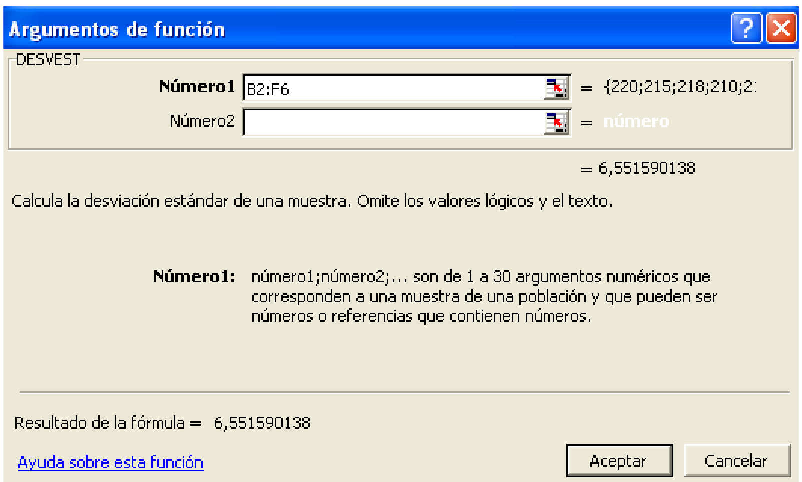
DESVESTP: Calcula la desviación estándar de todos los datos de una población.

Formato: DESVESTP(número1;número2;...)
Categoría: Estadísticas

Tomemos los datos del **ejemplo 5.2.1** para aplicar la fórmula de desviación estándar para datos muestrales. Copie los datos a una hoja en blanco en Excel:

	A	B	C	D	E	F	G
1							
2		220	215	218	210	210	
3		219	208	207	213	225	
4		213	204	225	211	221	
5		218	200	205	220	215	
6		217	209	207	211	218	
7							
8							

En la celda B8 active la función DESVEST, marcando en la primera casilla, losdatos del ejercicio y luego pulsando en el botón aceptar.



El resultado es de aproximadamente 6,5516.

	A	B	C	D	E	F	G
1							
2		220	215	218	210	210	
3		219	208	207	213	225	
4		213	204	225	211	221	
5		218	200	205	220	215	
6		217	209	207	211	218	
7							
8		6,5516					

Para datos agrupados, calcularemos la varianza tal cual como se mostró en la **sección 5.2.3** para luego calcular su raíz cuadrada con la función RAIZ:

fx

RAIZ: Calcula la raíz cuadrada de un número.

Formato: RAIZ(número1)
Categoría: Matemáticas y trigonométricas

Calculemos la raíz cuadrada de una tabla de frecuencia sencilla.

Ni	Clase	f
1	4	15
2	5	10
3	6	1
4	7	4
5	8	5
6	9	12
Total		47

En la celda B11 hallamos la media aritmética de la tabla.

	A	B	C	D	E
1					
2		**Ni**	**Clase**	**f**	
3		1	4	15	
4		2	5	10	
5		3	6	1	
6		4	7	4	
7		5	8	5	
8		6	9	12	
9		**Total**		47	
10					
11		6,2128			
12					

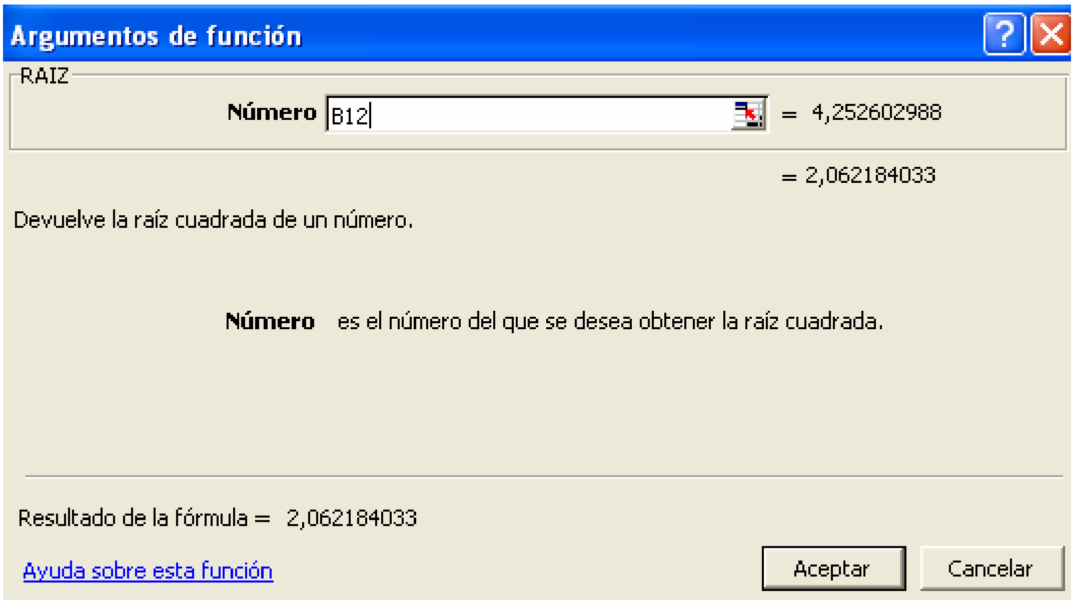
En una columna nueva colocamos las distancias de las clases respecto a la media, multiplicadas por sus frecuencias respectivas.

	A	B	C	D	E	F
1						
2		**Ni**	**Clase**	**f**	**Distancia**	
3		1	4	15	73,44	
4		2	5	10	14,71	
5		3	6	1	0,05	
6		4	7	4	2,48	
7		5	8	5	15,97	
8		6	9	12	93,22	
9		**Total**		47	199,87	
10						
11		6,2128				
12						
13						

Dividimos el total de las distancias al cuadrado por el número de datos (colocamos el resultado en la celda B12).

	A	B	C	D	E	F
1						
2		**Ni**	**Clase**	**f**	**Distancia**	
3		1	4	15	73,44	
4		2	5	10	14,71	
5		3	6	1	0,05	
6		4	7	4	2,48	
7		5	8	5	15,97	
8		6	9	12	93,22	
9		**Total**		47	199,87	
10						
11		6,2128				
12		4,2526				
13						

La desviación será igual a la raíz cuadrada del valor contenido en la celda B12.



La desviación estándar es de 2,0622.

	A	B	C	D	E	F
1						
2		**Ni**	**Clase**	**f**	**Distancia**	
3		1	4	15	73,44	
4		2	5	10	14,71	
5		3	6	1	0,05	
6		4	7	4	2,48	
7		5	8	5	15,97	
8		6	9	12	93,22	
9		**Total**		47	199,87	
10						
11		6,2128				
12		4,2526				
13		2,0622		=RAIZ(B12)		

5.4 COEFICIENTE DE VARIACIÓN

El coeficiente de variación permite comparar la dispersión entre dos poblaciones distintas e incluso, comparar la variación producto de dos variables diferentes (que pueden provenir de una misma población).

Estas variables podrían tener unidades diferentes, por ejemplo, podremos determinar si los datos tomados al medir el volumen de llenado de un embase de cierto líquido varían más que los datos tomados al medir la temperatura de el liquido contenido en el embase al salir al consumidor. El volumen los mediremos en centímetros cúbicos y la temperatura en grados centígrados.

El coeficiente de variación elimina la dimensionalidad de las variables y tiene en cuenta la proporción existente entre una medida de tendencia y la desviación típica o estándar.

Coeficiente de variación (Cv): Equivale a la razón entre la media aritmética y la desviación típica o estándar.
.

$$Cv = \frac{S}{\overline{X}} \text{ o } Cv = \frac{\sigma}{\mu}$$

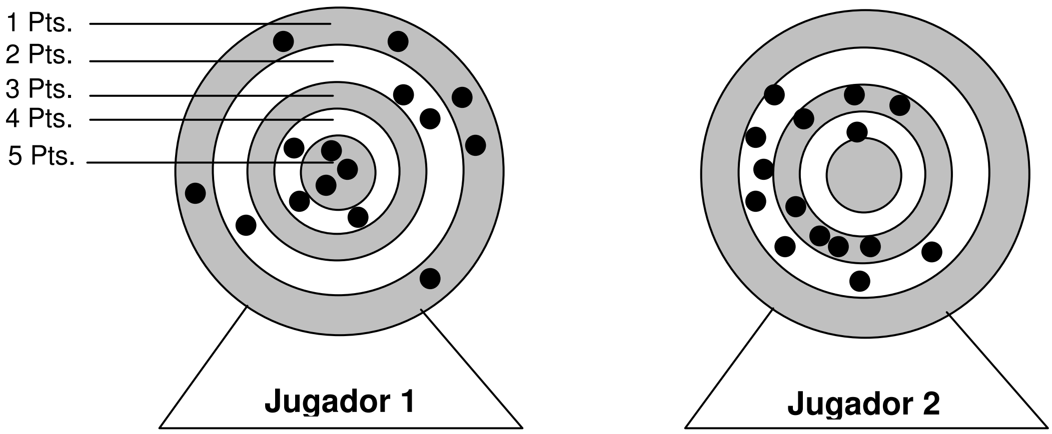
Si envés de la media aritmética se emplea la mediana, obtendremos el coeficiente de variación mediana.

$$Cv = \frac{S}{Me} \text{ o } Cv = \frac{\sigma}{Me}$$

Este índice solo se debe calcular para variables con todo los valores positivos, para dar seguridad de un $\overline{X}$ o μ mayores a cero (un coeficiente de variación positivo).

5.4.1 Ejemplo: Desviación estándar para datos no agrupados

En un juego de tiro al blanco con escopeta de perdigones por dos participantes a un tablero, obtienen el siguiente registro después de 15 disparos cada uno. Determinar el coeficiente de variación para ambos casos.



Disparo	f
1	6
2	3
3	0
4	3
5	3

Disparo	f
1	0
2	7
3	7
4	1
5	0

PASO 1: Calcular las medias aritméticas:

$$\overline{X}_1 = \frac{1x6 + 2x3 + 3x0 + 4x3 + 5x3}{15} = \frac{39}{15} = 2{,}6$$

$$\overline{X}_2 = \frac{1x0 + 2x7 + 3x7 + 4x1 + 5x0}{15} = \frac{39}{15} = 2{,}6$$

PASO 2: Calcular las varianzas

En este punto, la varianza es identificada por S^2.

$$S_1^2 = \frac{(1-2{,}6)^2.6 + (2-2{,}6)^2.3 + (3-2{,}6)^2.0 + (4-2{,}6)^2.3 + (5-2{,}6)^2.3}{15-1}$$

$$S_1^2 = \frac{39{,}6}{14} = 2{,}8286$$

$$S_2^2 = \frac{(1-2{,}6)^2.0 + (2-2{,}6)^2.7 + (3-2{,}6)^2.7 + (4-2{,}6)^2.1 + (5-2{,}6)^2.0}{15-1}$$

$$S_1^2 = \frac{5{,}6}{14} = 0{,}4$$

PASO 3: Calcular la desviación estándar a partir de la raíz cuadrada de la varianza.

$$S_1 = \sqrt{2{,}8286} = 1{,}6818$$

$$S_2 = \sqrt{0{,}4} = 0{,}6325$$

La puntuación de los disparos se aleja en promedio de la media aritmética en aproximadamente 1,6818 para el jugador 1 y 0,6325 para el jugador 2.

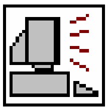
PASO 4: Calcular el coeficiente de variación.

$$Cv_1 = \frac{S_1}{\overline{X}_1} = \frac{1{,}6818}{2{,}6} = 0{,}6469$$

$$Cv_2 = \frac{S_2}{\overline{X}_2} = \frac{0{,}6325}{2{,}6} = 0{,}2433$$

El menor coeficiente de variación indica que el jugador 2 presento una dispersión menor de sus puntuaciones respecto a la media, caso contrario al jugador 1 donde la dispersión fue mayor.

5.4.2 Calculo del coeficiente de variación en Excel



Para calcular el coeficiente de variación con ayuda de Excel, debemos calcular primero la media aritmética y la desviación estándar. Por ejemplo, calculemos el coeficiente de variación para los siguientes datos:

	A	B	C	D	E	F	G
1							
2		1	1	1	1	1	
3		1	2	2	2	4	
4		4	4	5	5	5	

Empleando las fórmulas vistas en Excel, se halla la media y desviación (tomando los valores como muestrales):

	A	B	C	D	E	F	G
1							
2		1	1	1	1	1	
3		1	2	2	2	4	
4		4	4	5	5	5	
5							
6		**Media**	2,6	← =PROMEDIO(B2:F4)			
7		**Desviación**	1,6818	← =DESVEST(B2:F4)			
8							

El coeficiente de variación es el resulta de la división entre la desviación (C7) y la media (C6):

	A	B	C	D	E	F	G
1							
2		1	1	1	1	1	
3		1	2	2	2	4	
4		4	4	5	5	5	
5							
6		**Media**	2,6				
7		**Desviación**	1,6818				
8		**Cv**	0,6469	← =C7/C6			

5.5 EJERCICIOS PROPUESTOS

5.5.1 Calcular la desviación media a partir del registro de las siguientes edades de una muestra de 36 personas.

48	19	12	1	19	15
1	5	2	14	18	49
20	30	2	17	46	40
43	45	8	28	23	50
7	44	11	18	21	13
48	3	35	41	49	23

5.5.2 Calcular la desviación media a partir de la siguiente tabla de frecuencia.

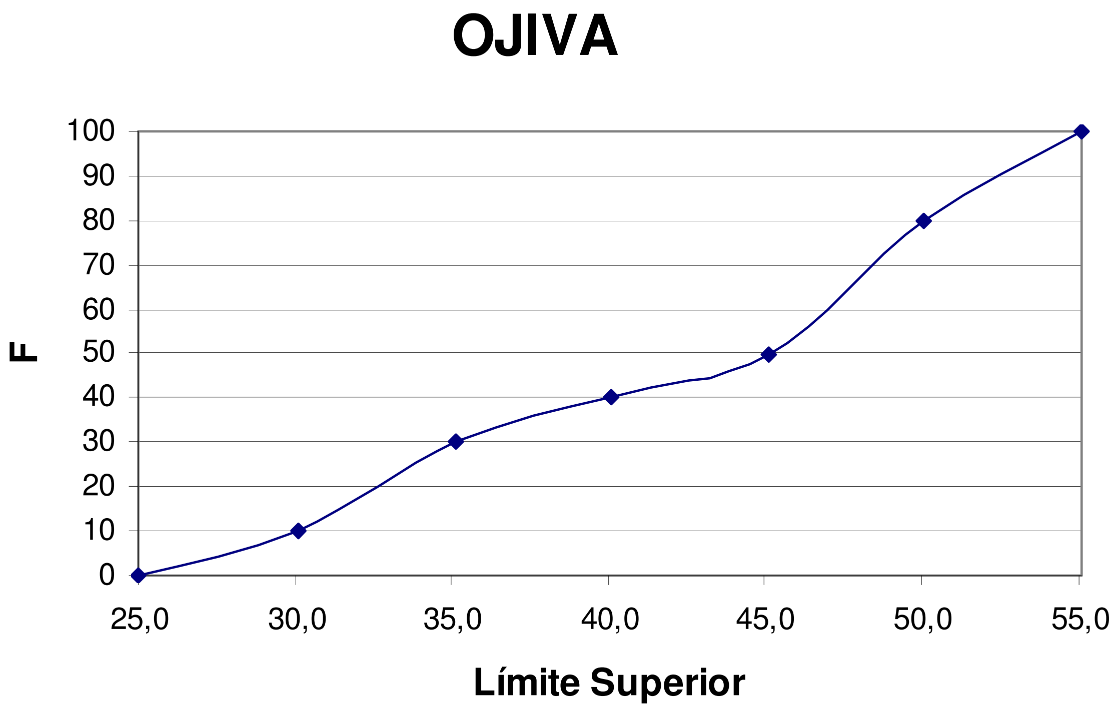
Nc	Lm	Ls	f
1	100,00	150,51	5
2	150,51	201,01	8
3	201,01	251,51	9
4	251,51	302,01	15
5	302,01	352,51	2
6	352,51	403,01	6
7	403,01	453,50	4
Total			**49**

5.5.3 Calcular la desviación media, varianza y desviación estándar a los datos mostrados en los **ejercicios 4.4.1, 4.4.2, 4.4.3 y 4.4.4**

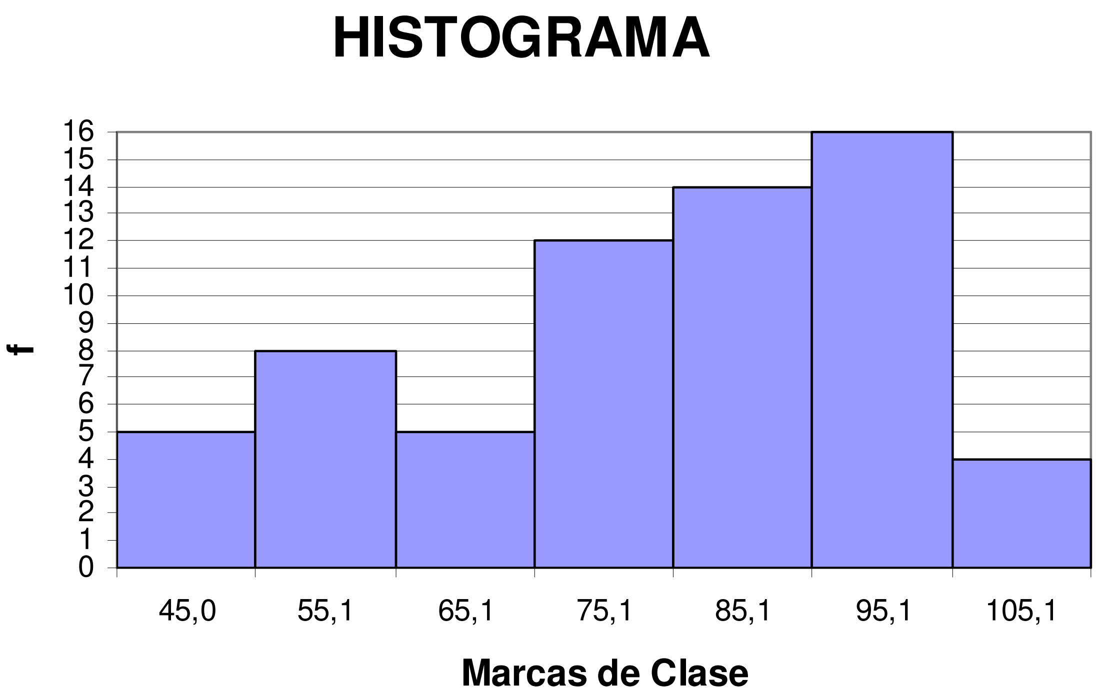
5.5.4 Calcule la desviación media, varianza y desviación estándar a partir de los siguientes datos sin agrupar y agrupándolos en una tabla de frecuencia tipo B (notar la variación de las medidas de dispersión en ambos casos).

49,15	46,17	53,28	49,41	49,00	36,14	41,65	51,75
45,13	43,00	41,95	45,95	52,66	47,50	37,43	48,53
47,24	47,55	51,17	52,69	37,12	49,39	35,20	45,14
35,20	40,59	54,06	47,05	47,04	53,13	53,88	42,33
45,16	35,87	35,02	39,33	48,64	51,83	49,89	36,13

5.5.5 Calcule la desviación media, varianza y desviación estándar a partir del siguiente gráfico de ojiva (dado en el **ejercicio 4.4.6**):



5.5.6 Calcule la desviación media, varianza y desviación estándar a partir del siguiente gráfico de ojiva (dado en el **ejercicio 4.4.7**):



5.6 CASO: EL RÍO MAGDALENA

El cuadro que figura más abajo da los caudales mensuales del rio Magdalena observados durante los meses del abril y mayo, desde 1988 a 2005 (la unidad de medida no se precisa).

AÑO	***ABRIL***	***MAYO***
1988	600	512
1989	227	211
1990	487	469
1991	560	370
1992	521	363
1993	423	272
1994	307	241
1995	390	253
1996	364	408
1997	284	233
1998	415	245
1999	255	199
2000	209	215
2001	230	297
2002	424	309
2003	528	303
2004	258	196
2005	242	166

Se desea ordenar estos datos y efectuar el análisis siguiente:

1. Dar una representación global de los caudales de abril y mayo. Graficar mediante dos histogramas los datos resumidos (***recomendación:*** agrupe los datos empleando tablas de frecuencia con iguales intervalos de clase).
2. Calcular la media de los caudales de abril, y la media de los caudales de mayo.
3. Calcular la desviación típica de los caudales de abril, y la desviación típica de los caudales de mayo.
4. Comparar los caudales de abril con los caudales de mayo, a partir de la información suministrada en la segunda y tercera pregunta.
5. Realizar conclusiones sobre: media, mediana, moda, frecuencias, desviaciones e histogramas de frecuencia.