MEDIDAS DE DISPERSIÓN CON EXCEL

DESVIACIÓN MEDIA O DESVIACIÓN PROMEDIO

La desviación media o desviación promedio es la media aritmética de los valores absolutos de las desviaciones respecto a la media aritmética.

a) Para Datos No Agrupados

Se emplea la fórmula:

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

Ejemplo ilustrativo: Calcular la desviación media de la distribución: 3, 8, 8, 8, 9, 9, 9, 18

Solución: Se calcula la media aritmética.

$$\bar{x} = \frac{\sum x_i}{n} = \frac{3+8+8+8+9+9+9+18}{8} = 9$$

Se calcula la desviación media.

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

$$\mathrm{DM} = \frac{|3-9| + |8-9| + |8-9| + |8-9| + |9-9| + |9-9| + |9-9| + |18-9|}{8}$$

$$\mathrm{DM} = \frac{6+1+1+1+0+0+0+9}{8} = \frac{18}{8} = 2{,}25$$

Empleando Excel:

	A	B	C
1	3		
2	8		
3	8		
4	8		
5	9		
6	9		
7	9		
8	18		
9	DM	2,25	=DESVPROM(A1:A8)

b) Para Datos Agrupados en Tablas de Frecuencia

Se emplea la fórmula:

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

Ejemplo ilustrativo: Calcular la desviación media en base a la siguiente tabla sobre las calificaciones de un estudiante en 12 asignaturas evaluadas sobre 10**.**

Calificación	Cantidad de asignaturas
6	4
7	2
8	3
9	2
10	1
Total	12

Solución:
Se calcula la media aritmética.

$$\bar{x} = \frac{\sum fx}{n} = \frac{4 \cdot 6 + 2 \cdot 7 + 3 \cdot 8 + 2 \cdot 9 + 1 \cdot 10}{12} = \frac{24 + 14 + 24 + 18 + 10}{12} = \frac{90}{12} = 7{,}5$$

Se llena la siguiente tabla:

x	f	$\lvert x - \bar{x}\rvert$	$f\lvert x - \bar{x}\rvert$
6	4	1,5	6
7	2	0,5	1
8	3	0,5	1,5
9	2	1,5	3
10	1	2,5	2,5
Total	12		14

Se emplea la ecuación de la desviación media.

$$DM = \frac{\sum f|x - \bar{x}|}{n} = \frac{14}{12} = 1{,}167$$

c) Para Datos Agrupados en Intervalos

Se emplea la fórmula:

$$DM = \frac{\sum f|xm - \bar{x}|}{n}$$

Donde xm es la marca de clase**.**

Ejemplo ilustrativo: Calcular la desviación media de un curso de 40 estudiantes en la asignatura de Estadística en base a la siguiente tabla:

Calificación	Cantidad de estudiantes
2-4	6
4-6	8
6-8	16
8-10	10
Total	40

Solución: Para calcular la media aritmética se llena la siguiente tabla:

Intervalo	f	xm	$f \cdot xm$
2-4	6	3	18
4-6	8	5	40
6-8	16	7	112
8-10	10	9	90
Total	40		260

Calculando la media aritmética se obtiene:

$$\bar{x} = \frac{\sum f \cdot xm}{n} = \frac{260}{40} = 6{,}5$$

Para calcular la desviación media se llena la siguiente tabla:

Intervalo	f	xm	$\lvert xm - \bar{x} \rvert$	$f\lvert xm - \bar{x} \rvert$
2-4	6	3	3,5	21
4-6	8	5	2,5	12
6-8	16	7	0,5	8
8-10	10	9	2,5	25
Total	40			66

$$DM = \frac{66}{40} = 1{,}65$$

VARIANZA Y DESVIACIÓN ESTÁNDAR

El teorema de Chebyshev, de autoría del matemático ruso Pafnuty Lvovich Chebyshev, establece que para todo conjunto de datos, por lo menos $1 - 1/k^2$ de las observaciones están dentro de k desviaciones estándar de la media, en donde k es cualquier número mayor que 1. Este teorema se expresa de la siguiente manera:

$$1 - \frac{1}{k^2}$$

Así por ejemplo, si se forma una distribución de datos con k =3 desviaciones estándar por debajo de la media hasta 3 desviaciones estándar por encima de la media, entonces por lo menos

$$1 - \frac{1}{3^2} = 1 - \frac{1}{9} = \frac{9-1}{9} = \frac{8}{9} = 0{,}8889 = 88{,}89\%$$

Interpretación: El 88,89% de todas las observaciones estarán dentro ± 3 desviaciones de la media.

a) Para Datos No Agrupados

La varianza para una población se calcula con:

$$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N}$$

Donde:

$x_i = observaciones\ individuales\ de\ la\ población$

μ = media aritmética poblacional

$N = número\ de\ observaciones\ de\ la\ población$

La desviación estándar poblacional se calcula con:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum(x_i - \mu)^2}{N}}$$

La varianza de la muestra se calcula con:

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

Donde:

$x_i = observaciones\ individuales\ de\ la\ muestra$

$\bar{x} = media\ aritmética\ de\ la\ muestra$

$n = número\ de\ obsevaciones\ de\ la\ muestra$

La desviación estándar de una muestra se calculó con:

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

Ejemplo ilustrativo: Considere que los siguientes datos corresponden al sueldo de una población: $350, $400, $500, $700 y $1000
1) Calcular la desviación estándar.
2) ¿Cuál es el intervalo que está dentro de k = 2 desviaciones estándar de la media?. ¿Qué porcentaje de las observaciones se encuentran dentro de ese intervalo?

Solución:
1) Para la calcular la desviación estándar se sigue los siguientes pasos:
a) Se calcula la media aritmética.

$$\mu = \frac{\sum x_i}{N} = \frac{350 + 400 + 500 + 700 + 1000}{5} = \frac{2950}{5} = \$\ 590$$

b) Se aplica la respectiva fórmula para calcular la varianza

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N}$$
$$\sigma^2 = \frac{(350 - 590)^2 + (400 - 590)^2 + (500 - 590)^2 + (700 - 590)^2 + (1000 - 590)^2}{5}$$
$$\sigma^2 = \frac{57600 + 36100 + 8100 + 12100 + 168100}{5} = \frac{282000}{5} = \$^2\ 56400$$

c) Se calcula la desviación estándar.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\$^2 56400} = \$237{,}4868$$

En Excel:

	A	B	C	D
1	350			
2	400			
3	500			
4	700			
5	1000			
6	σ	237,49	=DESVESTP(A1:A5)	

2) Cálculo del intervalo de k = 2 desviaciones estándar de la media.

Se transportan 2 desviaciones estándar (2 x $ 237,4868) = $ 474,97 por encima y por debajo de la media $\mu = \$590$

Por lo tanto se tiene un intervalo desde $ 590 - $474,97 = $ 115,03 hasta $ 590 + $474,97 = $ 1064,97

Aplicando el Teorema de Chebyshev
$$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = \frac{4 - 1}{4} = \frac{3}{4} = 0{,}75 = 75\%$$

Interpretación: Se puede afirmar de que por lo menos el 75% los sueldos están entre $ 115,03 y $ 1064,97

b) Para Datos Agrupados en Tablas de Frecuencia

La varianza para una población se calcula con:

$$\sigma^2 = \frac{\sum f(x_i - \mu)^2}{N}$$

Donde:

f = frecuencia absoluta.

La desviación estándar poblacional se calcula con:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum f(x_i - \mu)^2}{N}}$$

La varianza de la muestra se calcula con:

$$s^2 = \frac{\sum f(x_i - \bar{x})^2}{n-1}$$

La desviación estándar de una muestra se calcula con:

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

Ejemplo ilustrativo: Calcular la desviación estándar de los siguientes datos correspondientes a una muestra.

Calificaciones	f
4	3
5	6
6	4
7	13
8	7
10	6
Total	39

Solución: Se llena la siguiente tabla:

Calificaciones	f	fx
4	3	12
5	6	30
6	4	24
7	13	91
8	7	56
10	6	60
Total	39	273

Se calcula la media aritmética.

$$\bar{x} = \frac{\sum f \cdot x_i}{n} = \frac{273}{39} = 7$$

Se llena la siguiente tabla:

Calificaciones	f	fx_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$f(x_i - \bar{x})^2$
4	3	12	-3	9	27
5	6	30	-2	4	24
6	4	24	-1	1	4
7	13	91	0	0	0
8	7	56	1	1	7
10	6	60	3	9	54
Total	39	273			116

Se calcula la desviación estándar

$$s = \sqrt{\frac{116}{39-1}} = \sqrt{\frac{116}{38}} = \sqrt{3{,}0526} = 1{,}747$$

c) Para Datos Agrupados en Intervalos

La varianza para una población se calcula con:

$$\sigma^2 = \frac{\sum f(xm_i - \mu)^2}{N}$$

Donde:

$f = frecuencia\ absoluta$; $xm = marca\ de\ clase$

La desviación estándar poblacional se calcula con:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{\sum f(xm_i - \mu)^2}{N}}$$

La varianza de la muestra se calcula con:

$$s^2 = \frac{\sum f(xm_i - \bar{x})^2}{n-1}$$

La desviación estándar de una muestra se calcula con:

$$s = \sqrt{s^2} = \sqrt{\frac{\sum f(xm_i - \bar{x})^2}{n-1}}$$

Ejemplo ilustrativo: Calcular la desviación estándar de los siguientes datos correspondientes a una muestra.

Intervalo	f
60-65	5
65-70	20
70-75	40
80-85	27
85-90	8
Total	100

Solución: Se llena la siguiente tabla:

Intervalo	f	xm	$f \cdot xm$
60-65	5	62,5	312,5
65-70	20	67,5	1350
70-75	40	72,5	2900
80-85	27	82,5	2227,5
85-90	8	87,5	700
Total	100		7490

Se calcula la media aritmética.

$$\bar{x} = \frac{\sum f \cdot xm}{n} = \frac{7490}{100} = 74{,}9$$

Se llena la siguiente tabla:

Intervalo	f	xm	$f \cdot xm$	$xm_i - \bar{x}$	$(xm_i - \bar{x})^2$	$f(xm_i - \bar{x})^2$
60-65	5	62,5	312,5	-12,4	153,76	768,8
65-70	20	67,5	1350	-7,4	54,76	1095,2
70-75	40	72,5	2900	-2,4	5,76	230,4
80-85	27	82,5	2227,5	7,6	57,76	1559,52
85-90	8	87,5	700	12,6	158,76	1270,08
Total	100		7490			4924

Se calcula la desviación estándar.

$$s = \sqrt{\frac{4924}{100-1}} = \sqrt{\frac{4924}{99}} = \sqrt{49{,}737} = 7{,}052$$

OTRAS MEDIDAS DE DISPERSIÓN

a) RANGO, AMPLITUD TOTAL O RECORRIDO

Dada una serie de valores $x_1, x_2, \ldots . x_n$, su recorrido es la diferencia aritmética entre el máximo y el mínimo de estos valores.

$$Re = xmáx - xmín$$

Ejemplo ilustrativo: Calcula el rango de las siguientes distribuciones:
1) 4, 6, 8, 10, 12, 14, 16
2) 5, 10, 13, 13, 14, 15, 17

Solución:
$Re_1 = 16 - 4 = 12$; $Re_2 = 17 - 5 = 12$

En Excel:

	A	B	C	D	E	F	G
1	4				5		
2	6				10		
3	8				13		
4	10				13		
5	12				14		
6	14				15		
7	16				17		
8							
9	Recorrido	12	=MAX(A1:A7)-MIN(A1:A7)		Recorrido	12	=MAX(E1:E7)-MIN(E1:E7)

Ambas series tienen rango 12, pero están desigualmente distribuidas, pues mientras la primera se distribuye uniformemente a lo largo de todo el recorrido, la segunda tiene una mayor concentración en el centro.

La amplitud es una medida de dispersión cuya ventaja es la facilidad con que se calcula. Tiene en cambio las siguientes desventajas:

- En su cálculo sólo intervienen dos elementos del conjunto.

- Al aumentar el número de observaciones, puede esperarse que aumente la variabilidad. Puesto que la amplitud no tiene en cuenta el tamaño del conjunto, no es una medida adecuada para comparar la variabilidad de dos grupos de observaciones, a menos que éstos sean del mismo tamaño.

Nota: Cuando los datos están agrupados en intervalos se calcula la amplitud sacando la diferencia entre la marca de clase mayor y la marca de clase menor.

b) AMPLITUD INTERCUARTÍLICA

La amplitud intercuartílica es la distancia entre el tercer cuartil Q_3 y el primer cuartil Q_1.
Amplitud intercuartílica = tercer cuartil - primer cuartil = $Q_3 - Q_1$

c) RANGO SEMI-INTERCUARTIL O DESVIACIÓN CUARTÍLICA

La desviación cuartílica es la mitad de la distancia entre el tercer cuartil y el primero

$$DQ = \frac{Q_3 - Q_1}{2}$$

Ejemplo ilustrativo: Si el tercer cuartil = 24 y el primer cuartil = 10. ¿Cuál es la desviación cuartílica?

Solución: La amplitud intercuartílica es 24 - 10 = 14. Por lo tanto, la desviación cuartílica es:

$$DQ = \frac{14}{2}$$

d) RANGO PERCENTIL O AMPLITUD CUARTÍLICA

Cada conjunto de datos tiene 99 percentiles, que dividen el conjunto en 100 partes iguales.
La amplitud cuartílica es la distancia entre dos percentiles establecidos.

$$Rango\ percentil\ = P_{90} - P_{10}$$

DISPERSIÓN RELATIVA O COEFICIENTE DE VARIACIÓN

El Coeficiente de variación (CV) es una medida de la dispersión relativa de un conjunto de datos, que se obtiene dividiendo la desviación estándar del conjunto entre su media aritmética y se expresa generalmente en términos porcentuales. Para una población se emplea la siguiente fórmula:

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

Donde:

$CV = Coeficiente\ de\ variación;\ \sigma = desviación\ estándar\ de\ la\ población$

$\mu = media\ aritmética\ de\ la\ población$

Para una muestra se emplea la siguiente fórmula:

$$CV = \frac{s}{\bar{x}} \cdot 100\%$$

Donde:

$CV = Coeficiente\ de\ variación; s = desviación\ estándar\ de\ la\ muestra$

$\bar{x} = media\ aritmética\ de\ la\ muestra$

Ejemplo ilustrativo: Mathías, un estudiante universitario, tiene las siguientes calificaciones en las 10 asignaturas que recibe en su carrera: 8, 7, 10, 9, 8, 7, 8, 10, 9 y 10. Emily, una compañero de Mathías, tiene las siguientes calificaciones: 8, 9, 8, 7, 8, 9, 10, 7, 8 y 10. ¿Cuál estudiante tiene menor variabilidad en sus calificaciones?

Solución: Como se está tomando en cuenta todas las asignaturas, se debe calcular el coeficiente de variación poblacional.

En Excel:

	A	B	C	D	E	F
1	Mathias				Emily	
2	8				7	
3	7				7	
4	10				8	
5	9				8	
6	8				8	
7	7				8	
8	8				9	
9	10				9	
10	9				10	
11	10				10	
12	μ	8,60	=PROMEDIO(A2:A11)		8,40	=PROMEDIO(E2:E11)
13	σ	1,11	=DESVESTP(A2:A11)		1,02	=DESVESTP(E2:E11)
14	CV	12,95	=(B13/B12)*100		12,14	=(E13/E12)*100

Agrupando los datos en tablas de frecuencias:
Para Mathías se obtiene:
En Excel:

	A	B	C	D	E	F
1	x_i	f	$f(x_i-\mu)^2$			
2	7	2	5,12	=B2*(A2-B7)^2		
3	8	3	1,08	=B3*(A3-B7)^2		
4	9	2	0,32	=B4*(A4-B7)^2		
5	10	3	5,88	=B5*(A5-B7)^2		
6	Total	10	12,4	=SUMA(C2:C5)		
7	μ	8,6	=SUMAPRODUCTO(A2:A5;B2:B5)/SUMA(B2:B5)			
8						
9	σ	1,114	=RAIZ(C6/SUMA(B2:B5))			
10						
11	CV	12,948	=(B9/B7)*100			

Para Emily se obtiene:
En Excel:

	A	B	C	D	E	F
1	x_i	f	$f(x_i-\mu)^2$			
2	7	2	3,92	=B2*(A2-B7)^2		
3	8	4	0,64	=B3*(A3-B7)^2		
4	9	2	0,72	=B4*(A4-B7)^2		
5	10	2	5,12	=B5*(A5-B7)^2		
6	Total	10	10,4	=SUMA(C2:C5)		
7	μ	8,4	=SUMAPRODUCTO(A2:A5;B2:B5)/SUMA(B2:B5)			
8						
9	σ	1,020	=RAIZ(C6/SUMA(B2:B5))			
10						
11	CV	12,141	=(B9/B7)*100			

Interpretación: Por lo tanto, Emily tiene menor variabilidad en sus calificaciones

Fuente:

Suárez, Mario. & Tapia, Fausto. (2014). *Interaprendizaje de Estadística Básica*. Ibarra, Ecuador: Universidad Técnica de Norte

Suárez, Mario. (2014). *Probabilidades y Estadística empleando las TIC. Ibarra*, Ecuador: Imprenta GRAFICOLOR

Libros y artículos del Mgs. Mario Suárez sobre Aritmética, Álgebra, Geometría, Trigonometría, Lógica Matemática, Probabilidades, Estadística Descriptiva, Estadística Inferencial, Cálculo Diferencial, Cálculo Integral, y Planificaciones Didácticas se encuentran publicados en:
http://es.scribd.com/mariosuarezibujes
http://repositorio.utn.edu.ec/handle/123456789/760
http://www.docentesinnovadores.net/Usuarios/Ver/29591