CAPÍTULO 4

MEDIDAS DE DISPERSIÓN

En el capítulo anterior se estudiaron las medidas de tendencia central, que son un indicador de cómo los datos se agrupan o concentran en una parte central del conjunto. Sin embargo, para una información completa de dicho conjunto de datos hace falta saber el comportamiento opuesto, es decir, de qué manera se dispersan o se alejan algunos datos de esa parte central.

Por ejemplo, al tomar las temperaturas en una región A durante diferentes épocas del año y a distintas horas del día, se registraron los datos que se muestran en la columna A; por su parte, las de otra región diferente B, son las de la columna B .

Al obtener la media, en ambos casos resultó que la temperatura promedio fue de 20.68º, cuya interpretación podría ser que en torno, al rededor o cerca a 20.68º fluctúan los demás valores.

Como puede verse, eso es bastante aproximado para los datos de la columna A, no así para los de la B. Los datos más alejados en A son 19.3º y 22º, que realmente están próximos a 20.68º; en cambio, los datos más alejados en B son -3º y 39º, que están muy distantes del promedio.

	A	B
	19.3º	- 3º
	20º	0º
	20.2º	6º
	20.4º	22º
	21º	31.5º
	21.3º	34º
	21.3º	36º
	22º	39º
Promedio:	20.68º	20.68º

¿Por qué si en ambos casos se tiene igual promedio, no se puede afirmar lo mismo de los valores que están a su alrededor? La respuesta está en que no se ha tomado en cuenta la dispersión, es decir, la manera en que se disgregan los datos respecto de la media, pues en A casi no se dispersan mientras que en B sí, .Cabría decir que el conjunto de datos A es bastante compacto mientras que el B es muy dilatado.

Las principales medidas de dispersión son tres: El *rango*, la *desviación media* y la *desviación estándar*. De manera semejante a las medidas de tendencia central, las medidas de dispersión deben considerarse en sus dos opciones: cuando no están agrupados los datos y cuando están por intervalos.

4.1 EL RANGO

El *rango* o recorrido es la diferencia entre los datos mayor y menor del conjunto.

En un conjunto de datos, mientras mayor sea el rango, mayor será su dispersión y, a la inversa, mientras menor sea su rango, menor su dispersión. Dicho de otra forma; mientras mayor sea el rango, mayor “espacio” tendrán los datos para dispersarse, o mientras menor sea el rango, más estrechos estarán.

En los casos de las temperaturas del ejemplo anterior, el rango de A es $R = 22 - 19.3 = 2.7$, en cambio, el de B es $R = 39 - (-3) = 42$.

4.2 LA DESVIACIÓN MEDIA

Dado un conjunto de datos cuya media aritmética o promedio es $\overline{x}$, la diferencia o la distancia de cada valor nominal x a la media aritmética se llama desviación del dato x con respecto a la media $\overline{x}$. Es decir, es una medición de cuánto se alejó cada valor nominal x de la media.

x	desviación de x
50	50 - 75 = - 25
60	60 - 75 = - 15
70	70 - 75 = - 5
80	80 - 75 = 5
90	90 - 75 = 15
100	100 - 75 = 25

$\sum x = 450$

Por ejemplo, de los datos mostrados en la tabla de la derecha, la media aritmética es

$$\overline{x} = \frac{450}{6} = 75$$

entonces

La desviación del dato $x = 50$ con respecto de la media es $d = 50 - 75 = -25$. Lo mismo puede decirse de los demás datos.

Resulta obvio que siendo la media aritmética $\overline{x}$ el punto central de todos los valores de los datos x, existan simétricamente valores positivos y negativos, o lo que es lo mismo, la suma de todas las desviaciones a la media siempre es cero. Para evitar lo anterior, dicha suma se toma como valor absoluto, esto es

La desviación d del dato $x = 50$ con respecto de la media $\overline{x}$ es $d = |50 - 75| = 25$

La desviación d del dato $x = 60$ con respecto de la media $\overline{x}$ es $d = |60 - 75| = 15$

La desviación d del dato $x = 70$ con respecto de la media $\overline{x}$ es $d = |70 - 75| = 5$

y así sucesivamente.

4.2.1 LA DESVIACIÓN MEDIA PARA FRECUENCIAS SIMPLES

Cuando los datos recolectados han sido organizados en una tabla de frecuencias simples, es decir, sin agrupar, la desviación media DM se calcula por medio de la fórmula:

$$\boxed{DM = \frac{\sum f\,|x - \overline{x}|}{n}}$$

Por lo tanto, deben añadirse a la tabla original tres columnas: la primera encabezada con fx, que servirá para calcular la media aritmética; la segunda encabezada con $|x - \overline{x}|$, que servirá para obtener la tercera, $f\,|x - \overline{x}|$ con la que se obtiene el numerador de la fórmula luego de realizar su sumatoria.

Ejemplo 1: Obtener la desviación media del conjunto de datos mostrado en la tabla de la derecha.

edad x	f
45	2
46	1
47	3
48	3
49	5
50	6
51	2
52	4
53	2
	28

Solución: La tabla de la derecha es la original a la que deben agregársele tres columnas:

a) la primera columna agregada se encabeza con fx que representa la multiplicación de cada frecuencia f por su respectivo valor nominal x.

Al concluir de llenar esta columna se debe efectuar la sumatoria $\sum fx$ para calcular la media aritmética. Ver la tabla completa en la siguiente página.

b) la segunda columna agregada se encabeza con $\left|x-\overline{x}\right|$, que representa el valor absoluto de la resta de cada valor nominal x menos la media $\overline{x}$ obtenida en el paso anterior, y

c) la tercera columna agregada se encabeza con $f\left|x-\overline{x}\right|$, que representa la multiplicación de cada frecuencia f (2ª columna) por el valor absoluto correspondiente obtenido en la 4ª columna.

Al concluir de llenar esta columna se debe efectuar la sumatoria $\sum f\left|x-\overline{x}\right|$.

La tabla, con esas columnas agregadas, queda así:

x	f	fx	$\left\|x-\overline{x}\right\|$	$f\left\|x-\overline{x}\right\|$
45	2	90	4.357	8.714
46	1	46	3.357	3.357
47	3	141	2.357	7.071
48	3	144	1.357	4.071
49	5	245	0.357	1.785
50	6	300	0.643	3.858

x	f	fx	$\left\lvert x-\overline{x}\right\rvert$	$f\left\lvert x-\overline{x}\right\rvert$
51	2	102	1.643	3.286
52	4	208	2.643	10.572
53	2	106	3.643	7.286
	28	1382		50

$$\overline{x} = \frac{\sum fx}{\sum f}$$

$$\overline{x} = \frac{1382}{28} = 49.357$$

Por lo tanto

$$DM = \frac{\sum f\left\lvert x-\overline{x}\right\rvert}{\sum f}$$

$$DM = \frac{50}{28} = 1.785$$

Esto significa que el promedio de alejamiento de todos los valores respecto de la media $\overline{x} = 49.357$ es de 1.785.

C U E S T I O N A R I O 4.1

1) Obtener la *desviación media* de los datos organizados en el cuestionario 2.3.

4.2.2 LA DESVIACIÓN MEDIA CON TABLAS POR INTERVALOS

Cuando los datos han sido organizados en clases o intervalos, la *desviación media* se obtiene de manera similar a los procesos anteriores, es decir, con la misma fórmula aplicada a la organización de frecuencias simples, solamente que ***x*** debe ser el punto medio del intervalo.

Esto significa que al conjunto de datos original deben añadirse a la tabla ahora cuatro columnas: la primera encabezada con *x* para señalar el punto medio de cada intervalo; la segunda encabezada con *fx* que servirá para calcular la media aritmética; la tercera encabezada con $|x - \overline{x}|$ que servirá para obtener la cuarta, y la cuarta con $f|x - \overline{x}|$ que servirá para obtener el numerador de la fórmula luego de realizar su sumatoria.

Ejemplo 1: Cien datos recolectados se organizaron en siete intervalos, los que se muestran en la tabla de la derecha. Obtener la desviación media.

intervalo	*f*
4 - 9	12
10 - 15	11
16 - 21	13
22 - 27	19
28 - 33	21
34 - 39	16
40 - 45	8
	100

Solución: La tabla de la derecha es la original a la que deben agregársele cuatro columnas:

a) la primera columna agregada se encabeza con *x* que representa el punto medio de cada intervalo. Ver la tabla completa en la siguiente página.

b) la segunda columna agregada se encabeza con *fx*.

Al concluir de llenar esta columna se debe efectuar la sumatoria $\sum fx$ para calcular la media aritmética:

$$\overline{x} = \frac{\sum fx}{\sum f}$$

$$\overline{x} = \frac{2486}{100} = 24.86$$

c) la tercera columna agregada se encabeza con $\left| x - \overline{x} \right|$ que representa el valor absoluto de la resta de cada punto medio x del intervalo menos la media $\overline{x}$ obtenida en el paso anterior. Así:

primera fila: $\left| x - \overline{x} \right| = \left| 6.5 - 24.86 \right| = 18.36$

segunda fila: $\left| x - \overline{x} \right| = \left| 12.5 - 24.86 \right| = 12.36$

tercera fila: $\left| x - \overline{x} \right| = \left| 18.5 - 24.86 \right| = 6.36$ etc.

d) la cuarta columna agregada se encabeza con $f\left| x - \overline{x} \right|$ que representa la multiplicación de cada frecuencia f (2ª columna) por el valor absoluto correspondiente obtenido en la 5ª columna. Al concluir de llenar esta columna se debe efectuar la sumatoria $\sum f\left| x - \overline{x} \right|$.

La tabla con esas columnas agregadas queda así:

intervalo	f	x	fx	$\left\| x - \overline{x} \right\|$	$f\left\| x - \overline{x} \right\|$
4 - 9	12	6.5	78	18.36	220.32
10 - 15	11	12.5	137.5	12.36	135.96
16 - 21	13	18.5	240.5	6.36	82.68
22 - 27	19	24.5	465.5	0.36	6.84
28 - 33	21	30.5	640.5	5.64	118.44
34 - 39	16	36.5	584	11.64	186.24
40 - 45	8	42.5	340	17.64	141.12
	100		2486		891.6

$$DM = \frac{\sum f\left| x - \overline{x} \right|}{n}$$

$$DM = \frac{891.6}{100} = 8.916$$

Esto significa que el promedio de alejamiento de todos los valores respecto de la media, es de 8.916.

CUESTIONARIO 4.2

1) Obtener la *desviación media* de los datos organizados en el cuestionario 2.3, página 16.

4.3 LA DESVIACIÓN ESTÁNDAR

La tercera medida de dispersión se llama *desviación estándar*, porque con ella se pueden estandarizar en todos los casos, todas las desviaciones de datos recolectados, como se verá más adelante.

La *desviación estándar* se simboliza con la letra griega σ si se trata de una población y con la letra s si se trata de una muestra.

4.3.1 LA DESVIACIÓN ESTÁNDAR PARA FRECUENCIAS SIMPLES

Cuando los datos están ordenados en una distribución de frecuencias simples, la desviación estándar para una población se calcula mediante la fórmula

$$\sigma = \sqrt{\frac{\sum f\left(x-\overline{x}\right)^2}{n}}$$

en donde σ = desviación estándar de la población
f = frecuencia
x = valor nominal
$\overline{x}$ = media aritmética

Cuando los datos están ordenados en una distribución de frecuencias simples, la desviación estándar para una muestra se calcula mediante la fórmula

$$s = \sqrt{\frac{\sum f\left(x-\overline{x}\right)^2}{n-1}}$$

en donde s = desviación estándar de la muestra
f = frecuencia
x = valor nominal
$\overline{x}$ = media aritmética

Ejemplo 1: Obtener la desviación estándar del conjunto de datos de la muestra de la siguiente tabla.

Solución: La tabla de la derecha es la original a la que deben agregársele cuatro columnas:

a) la primera columna agregada se encabeza con fx que representa la multiplicación de cada frecuencia f por su

x	f
45	2
46	1
47	3
48	3
49	5
50	6
51	2
52	4
53	2
	28

respectivo valor nominal x. Al concluir de llenar esta columna se debe efectuar la sumatoria $\sum fx$ para calcular la media aritmética:

$$\overline{x} = \frac{\sum fx}{n}$$

$$\overline{x} = \frac{1382}{28} = 49.357$$

(ver tabla completa de la siguiente página)

b) la segunda columna agregada se encabeza con $\left(x - \overline{x}\right)$ que representa la resta de cada valor nominal x menos la media $\overline{x}$ obtenida en el paso anterior. Esta columna es opcional, pues directamente se puede elevar al cuadrado y el respectivo valor vaciarlo en la columna que se especifica en el siguiente inciso; de la siguiente forma:

primera fila:	$x - \overline{x}$	=	45 - 49.3571	=	- 4.3571
segunda fila:	$x - \overline{x}$	=	46 - 49.3571	=	- 3.3571
tercera fila:	$x - \overline{x}$	=	47 - 49.3571	=	- 2.3571
cuarta fila:	$x - \overline{x}$	=	48 - 49.3571	=	- 1.3571

y así sucesivamente.

c) la tercera columna agregada, o segunda en caso de haber omitido la anterior, se encabeza con $\left(x - \overline{x}\right)^2$ que representa el cuadrado de cada valor obtenido en la columna anterior:

primera fila:	$\left(x - \overline{x}\right)^2$	=	$(-4.3571)^2$	=	18.9843
segunda fila:	$\left(x - \overline{x}\right)^2$	=	$(-3.3571)^2$	=	11.2701
tercera fila :	$\left(x - \overline{x}\right)^2$	=	$(-2.3571)^2$	=	5.5559
cuarta fila:	$\left(x - \overline{x}\right)^2$	=	$(-1.3571)^2$	=	1.8417

y así sucesivamente.

d) la siguiente columna agregada se encabeza con $f\left(x-\overline{x}\right)^2$ que representa el producto de cada frecuencia f por su correspondiente cuadrado obtenido en la columna anterior.

La tabla con esas columnas agregadas queda así:

x	f	fx	$\left(x-\overline{x}\right)^2$	$f\left(x-\overline{x}\right)^2$
45	2	90	18.9834	37.9668
46	1	46	11.2694	11.2694
47	3	141	5.5554	16.6662
48	3	144	1.8414	5.5242
49	5	245	0.1274	0.6370
50	6	300	0.4134	2.4804
51	2	102	2.6994	5.3988
52	4	208	6.9854	27.9416
53	2	106	13.2714	26.5428
	28	1382		134.4272

Se tiene con esta tabla toda la información requerida para utilizar la fórmula de la desviación estándar para una muestra:

$$s=\sqrt{\frac{\sum f\left(x-\overline{x}\right)^2}{n-1}}$$

$$s=\sqrt{\frac{134.4272}{28-1}}$$

$$\boxed{s=2.231}$$

Para calcular la desviación estándar de una muestra a veces puede resultar más simple emplear la siguiente fórmula:

$$s = \sqrt{\frac{\sum fx^2 - \frac{\left(\sum fx\right)^2}{n}}{n-1}}$$

Aplicándola a la tabla del ejemplo anterior, se tiene que

x	f	x^2	fx^2	fx
45	2	2025	4050	90
46	1	2116	2116	46
47	3	2209	6627	141
48	3	2304	6912	144
49	5	2401	12005	245
50	6	2500	15000	300
51	2	2601	5202	102
52	4	2704	10816	208
53	2	2809	5618	106
	28		68346	1382

$$s = \sqrt{\frac{68346 - \frac{(1382)^2}{28}}{28-1}}$$

$$s = \sqrt{\frac{68346 - 62.9851}{27}}$$

$$\boxed{s = 2.231}$$

CUESTIONARIO 4.3

1) Obtener la *desviación estándar* de los datos organizados en el cuestionario 2.3, considerándolos como muestras.

4.3.2 LA DESVIACIÓN ESTÁNDAR PARA FRECUENCIAS POR INTERVALOS

Cuando los datos han sido organizados en clases o intervalos, la *desviación estándar* se obtiene de manera similar a los procesos anteriores, es decir, con la misma fórmula aplicada a la organización de frecuencias simples, solamente que x debe ser el punto medio del intervalo.

Ejemplo 1: Cien datos recolectados en una muestra se organizaron en los siete intervalos de la siguiente tabla. Obtener la *desviación estándar*.

Solución: La tabla de la derecha es la original a la que deben agregársele cuatro columnas:

a) la primera columna agregada se encabeza con x que representa el punto medio de cada intervalo.

b) la segunda columna agregada se encabeza con fx que representa la multiplicación de cada frecuencia f por su respectivo punto medio x del intervalo. Al concluir de llenar esta columna se debe efectuar la sumatoria $\sum fx$ para calcular la media aritmética.

intervalo	*f*
4 - 9	12
10 - 15	11
16 - 21	13
22 - 27	19
28 - 33	21
34 - 39	16
40 - 45	8
	100

$$\overline{x} = \frac{\sum fx}{n}$$

$$\overline{x} = \frac{2486}{10} = 2.486$$

c) la tercera columna agregada se encabeza con $x - \overline{x}$ que representa la resta de cada punto medio x del intervalo menos la media $\overline{x}$ obtenida en el paso anterior. Aunque debe recordarse que esta columna es opcional si el estudiante puede sin equivocarse obtener directamente sus cuadrados.

d) la cuarta columna agregada se encabeza con $\left(x - \overline{x}\right)^2$ que representa los cuadrados de cada resta obtenidos en la columna anterior.

e) la quinta columna agregada se encabeza con $f\left(x - \overline{x}\right)^2$ en donde se vaciarán los resultados de cada producto de la frecuencia f por el respectivo valor de la columna anterior.

La tabla, con esas columnas agregadas, queda así:

intervalo	f	x	fx	$\left(x - \overline{x}\right)^2$	$f\left(x - \overline{x}\right)^2$
4 - 9	12	6.5	78	337.0896	4 045.0752
10 - 15	11	12.5	137.5	152.7697	1 680.4656
16 - 21	13	18.5	240.5	40.4496	525.8448
22 - 27	19	24.5	465.5	0.1296	2.4624
28 - 33	21	30.5	640.5	31.8096	668.0016
34 - 39	16	36.5	584	135.4896	2 167.8336
40 - 45	8	42.5	340	311.1696	2 489.3568
	100		2 486		11 579.04

Sustituyendo en la fórmula de la desviación estándar

$$s = \sqrt{\frac{\sum f\left(x - \overline{x}\right)^2}{n - 1}}$$

$$s = \sqrt{\frac{11579.04}{100-1}}$$

$$\boxed{s = 10.8148}$$

CUESTIONARIO 4.4

1) Obtener la *desviación estándar* de los datos organizados en el cuestionario 2.3, tomándolos como muestras.