

CAPITULO 4

MEDIDAS DE DISPERSION

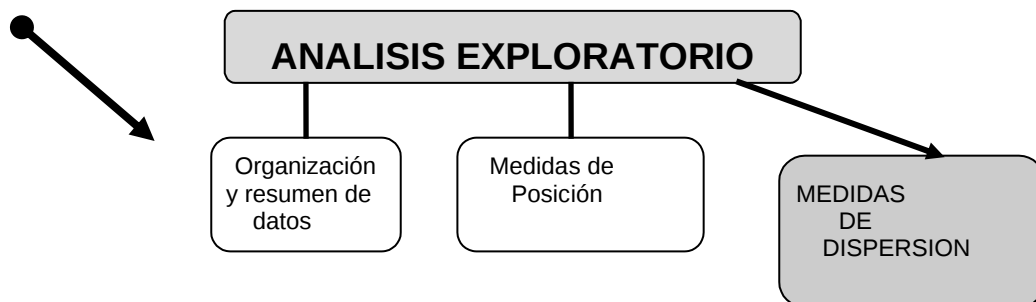
OBJETIVOS

- Identificar las medidas de dispersión más frecuentemente utilizadas para su aplicación en el tratamiento de los datos.
- Interpretar cada medida de dispersión para un completo análisis de la información relevada.

CONTENIDOS

1. Introducción.....
2. Medidas de dispersión.....
 - 2.1. Recorrido o rango.....
 - 2.2. Recorrido intercuartílico.....
 - 2.3. Desviación media absoluta.....
 - 2.4. Desviación mediana absoluta.....
 - 2.5. La varianza y el desvío estándar.....
 - 2.6. Teorema de Tchebyshev y Regla empírica.....
 - 2.7. Coeficiente de variación.....
3. Asimetría y curtosis
4. Indicadores de control.....
5. Ejercicios propuestos.....

UD. ESTÁ AQUÍ



1. Introducción

En el capítulo anterior vimos que estaba dedicado a las medidas de posición. Igualmente es importante describir la forma en que las observaciones están **dispersas, distribuidas o diseminadas**, a cada lado del centro de la distribución de las mismas. A esto por lo general se le conoce como **dispersión, variación o variabilidad de los datos** y permiten evaluar la confiabilidad o representatividad de la medida de tendencia central, sobre todo la media aritmética. En este capítulo se

exploran las medidas comúnmente empleadas para describir la dispersión de una muestra o población de observaciones.

Las medidas de dispersión son importantes debido a que dos muestras de observaciones con un mismo valor central pueden tener una variabilidad muy distinta. Por ejemplo, supongamos que se aplica una prueba idéntica a un grupo de 20 hombres y a un grupo de 20 mujeres. Las distribuciones de las calificaciones para los dos grupos son las siguientes:

Calificación de la prueba:	20	30	40	50	60	70	80	90	100
Hombres (n = 20)		2	4	7	6	1			
Mujeres (n = 20)	1	1	2	3	6	3	2	1	1

La calificación promedio para cada grupo es la misma (60 puntos), pero la dispersión de los valores para los dos grupos son diferentes. Los 20 hombres muestran muy poca variabilidad en sus calificaciones de prueba, mientras que las calificaciones de las mujeres muestran una variabilidad mucho mayor. Por lo tanto, las calificaciones de los hombres se hallan más concentradas alrededor de la calificación promedio, resultando más representativo para las personas de sexo masculino.

Las medidas de dispersión que veremos en este texto, son las siguientes:

- Rango o recorrido
- El rango intercuartílico
- Desviación media absoluta
- La varianza
- El desvío estándar
- La desviación media
- La desviación mediana
- El coeficiente de variación

2. Medidas de dispersión

2.1 Rango o recorrido

2.1.1 Concepto y cálculo

El rango está determinado por los dos valores extremos de los datos muestrales; es simplemente la diferencia entre el mayor y el menor valor de las observaciones.

$$R = X_{\text{máxima}} - X_{\text{mínima}}$$

Cuando hablamos del Rango, estamos afirmando alguna cosa del grado de concentración de los datos. Es evidente que cuanto mayor es el rango mayor es la dispersión o variabilidad de los valores de la variable.

Cuando tenemos los datos agrupados en intervalos de clase, el rango, es la diferencia entre el límite superior del último intervalo y el límite inferior del primer intervalo de clase.

El rango tiene el inconveniente de sólo tener en cuenta los dos valores extremos de la serie, sin considerar el conjunto de valores intermedios. Es un pobre indicador de la dispersión o variabilidad.

Ejemplo 5.1

Para los valores, supongamos que sean los pesos de algunos alumnos:

40, 45, 48, 52, 54, 62, y 70

Tenemos que el rango es:

$$R = 70 - 40 = 30 \text{ kg}$$

La interpretación sería: La diferencia entre el alumno que más pesa y el que menos pesa es de 30 kg.

El rango ofrece una medida deficiente de la dispersión, debido a que sólo emplea los valores extremos para su cálculo e ignora la naturaleza de la variación entre todas las demás observaciones. Por lo tanto está altamente influenciado por los valores extremos. Por ejemplo, podemos tener dos series de observaciones con el mismo valor del rango, pero su distribución es distinta. Lo veamos con un ejemplo:

1) 2, 2, 2, 3, 3, 8, 8, 10, 12, 20

2) 2, 3, 5, 6, 8, 9, 10, 14, 16, 20

En las dos series el rango es el mismo: $R = 20 - 2 = 18$. Pero podemos observar que en la segunda serie es mayor la dispersión.

La interpretación de tal medida es: la diferencia entre el mayor valor y el menor valor de la serie es de 18.

Una medida un poco más eficiente es el rango intercuartílico.

Actividad 1

Los siguientes datos corresponden a cuentas por cobrar (en miles de \$) de una empresa. Calcular el rango.

22 31 31.5 38 40 41 50.8 72

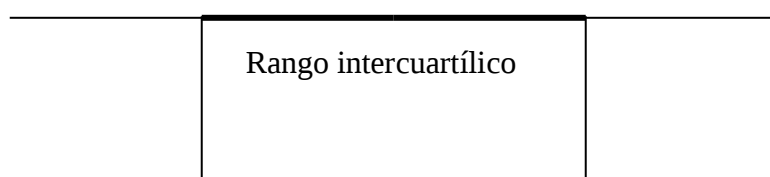
2.2 Rango intercuartílico

Es la diferencia entre el tercer cuartil y el primer cuartil. Esta medida evita el influjo que produce los valores extremos y considera los dos valores que dejan entre sí el 50% de los valores centrales de la variable.

$$RI = Q_3 - Q_1$$

A través del siguiente diagrama podemos observar el recorrido intercuartílico:

X



Q₁Q₃**Ejemplo 5.2**

$$\begin{array}{c|c} 1/4 & 1/4 \end{array}$$

De las dos series no tabuladas anteriores, calcular el rango intercuartílico.

1. 2, 2, 2, 3, 3, 8, 8, 10, 12, 20
2. 2, 3, 5, 6, 8, 9, 10, 14, 16, 20

Solución:

Dado que los datos no están tabulados, se opera de la siguiente forma:

Se ordenan los datos en forma ascendente y luego se localiza el punto de posición:

$$\frac{n+1}{4}$$

Si el resultado del cociente es un número entero, entonces Q₁ es igual al valor de la variable correspondiente al resultado del cociente. Si el mismo es igual a “dos”, entonces, no está indicando que el primer cuartil ocupa el segundo lugar de la serie de datos. Ahora si el resultado no es un número entero, se realiza una interpolación lineal entre los dos valores correspondientes a las dos observaciones entre las cuales se encuentra la fracción. Si el resultado por ejemplo es 2.3, significa que el primer cuartil es el segundo dato más un 30% de la diferencia entre los valores de las observaciones segunda y tercera. Hecha estas aclaraciones nos aboquemos al ejemplo:

1.

$$\frac{n+1}{4}$$

2, 2, 2, 3, 3, 8, 8, 10, 12, 20

$$\frac{10+1}{4} = 2.75$$

$$Q_1 = 2 + (3 - 2)(0.75) = 2$$

$$\frac{3(10+1)}{4} = 8.25$$

$$Q_3 = 10 + (12 - 10)(0.25) = 10.5$$

$$RI = 10.5 - 2 = 8.5$$

Q₁
↓

Q₃
↓

2, 2 2, 3, 3, 8, 8, 10, 12, 20

Una pregunta que nos permitiría interpretar con más claridad el rango intercuartílico es: ¿Entre qué valores se encuentra el 50% central de los datos?. La respuesta sería la diferencia entre el tercer cuartil y el primer cuartil, de esta forma nos permite tener una idea donde se encuentra el 50% central de los valores que adopta la variable. Para poder afirmar algo del valor que nos dé el RI debemos compararlo con otro RI de otra muestra de la misma población.

2.

2, 3, 5, 6, 8, 9, 10, 14, 16, 20

$$\frac{10 + 1}{4} = 2.75$$

$$Q_1 = 3 + (2 - 3)(0.75) = 2.25$$

$$\frac{3(10 + 1)}{4} = 8.25$$

$$Q_3 = 14 + (14 - 16)(0.25) = 13.5$$

$$RI = 13.5 - 2.25 = 11.25$$

Conclusión: En la primera serie están más concentrados en el 50% central de los datos que en la segunda serie, donde se hallan más dispersos al ser mayor el rango intercuartílico. Si realizáramos un diagrama de puntos, observaríamos que los puntos centrales estarían más dispersos en la segunda serie.

Actividad 2

Se selecciona una muestra de 11 obreros de una fábrica, y se les pregunta las horas extras que realizaron en la semana:

2 4 4 5 5.5 6 6 6 6.5 3 3

- ✓ Calcular el rango intercuartílico.
- ✓ ¿Qué ocurre si uno de ellos se equivocó y se rectifica afirmando que hizo 12 horas? ¿El resultado o conclusión es igual?

2.3 Desviación media absoluta

La desviación media absoluta es la sumatoria de la diferencia entre los valores absolutos observados y la media aritmética de ellos divididos por el total de observaciones. Se simboliza como “DM”. Es una medida superior al rango debido a que

tiene en cuenta todas las observaciones. Cada valor de la variable es ponderado y nos indica cuán alejados, en promedio, se encuentran de la media.
Para datos no tabulados tenemos:

$$DM = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$

Para datos tabulados nos queda:

$$DM = \frac{\sum_{i=1}^n |x_m - \bar{x}| f_i}{n}$$

Cuando los datos no están tabulados debemos realizar la diferencia de cada valor de la variable con la media aritmética. A cada diferencia en otra columna tomamos el valor absoluto y luego hacemos la suma, de esa forma ya tenemos el numerador. Finalmente dividimos por el total de observaciones.

Si los datos están tabulados con o sin intervalos de clase, el paso adicional es multiplicar las diferencias absolutas por cada frecuencia absoluta cada valor que toma la variable.

Ejemplo 5.3

En una localidad de la provincia se aprobaron ocho proyectos del programa social “Componentes materiales Tipología VI”. El número de beneficiarios del Plan Jefes/as de Hogar por cada proyecto se especifica en orden ascendente:

1 2 2 3 3 3 4 4

Calcule la dispersión media absoluta.

Solución:

La media aritmética de los datos es 2.75.

Número de beneficiarios	Desviación $x - \bar{x}$	Desviación absoluta $ x - \bar{x} $
1	-1.75	1.75
2	-0.75	0.75
2	-0.75	0.75
3	0.25	0.25
3	0.25	0.25
3	0.25	0.25
4	1.25	1.25
4	1.25	1.25
Total	0	6.5

Hacemos los cálculos:

$$DM = \frac{6.5}{8} = 0.81$$

Vamos a obtener la dispersión media absoluta con datos agrupados en intervalos de clase. Para ello se calcula en primer lugar la media aritmética de los datos agrupados. Como estamos trabajando con datos agrupados en intervalos de clase, en otra columna, determinamos los desvíos de las marcas de clase con respecto a la media. Hallamos el valor absoluto de los desvíos calculados anteriormente y multiplicamos a cada uno por su respectiva frecuencia. Finalmente sumamos los productos y a esa suma la dividimos por el total de observaciones. En una tabla realizaremos los pasos mencionados a través de un ejemplo:

Ejemplo 5.4

Los datos que se ven en el siguiente cuadro agrupados en intervalos de clase corresponden a la cantidad de préstamos otorgados a pequeñas y medianas empresas, en un período determinado, por entidades financieras de una ciudad, el objetivo de los mismos fue la adquisición de bienes de capital y capital de trabajo. Calcule la dispersión media absoluta:

Nº de préstamos	Nº de empresas	Marca de clase	fi Xi	Desvíos $x - \bar{x}$	Desviación absoluta $ x - \bar{x} $	$ x - \bar{x} f_i$
1 – 5	17	3	51	-6.42	6.42	109.14
5 – 9	19	7	133	-2.42	2.42	45.98
9 – 13	13	11	143	1.58	1.58	20.54
13 – 17	9	15	135	5.58	5.58	50.22
17 – 21	6	19	114	9.58	9.58	57.48
21 – 25	2	23	46	13.58	13.58	27.16
N	66		622	*	39.16	310.52

$$\bar{x} = \frac{622}{66} = 9.42 \cong 9 \text{ préstamos}$$

$$DM = \frac{310.52}{66} = 4.71$$

Actividad 3

* Los desvíos en este caso particular no dan cero por el redondeo de decimal.

Las edades de los suscriptores de una revista especializada en cine, se observan en el siguiente cuadro. Calcular la desviación media absoluta.

Edades	Nº de suscriptores
15-20	23
20-25	47
25-30	92
30-35	105
35-40	250
40-45	240
45-50	190
50-55	156
55-60	120

2.4 Desviación mediana absoluta

La desviación mediana absoluta es el promedio de los valores absolutos de las diferencias entre cada dato y la mediana de éstos. Se simboliza como **Dme**.

Las fórmulas para datos no agrupados y agrupados son las siguientes:

Para datos no agrupados

$$DMe = \frac{\sum_{i=1}^n |x_i - Me|}{n}$$

Para datos agrupados

$$DMe = \frac{\sum_{i=1}^n |x_m - Me| f_i}{n}$$

Ejemplo 5.5

Los pesos de nueve alumnos son los que se indican a continuación:

12 14 15 18 19 20 21 24 25

Calcular la desviación mediana absoluta.

En este ejemplo la mediana es igual a 19.

Peso de los alumnos	Desviación $x - Me$	Desviación absoluta $ x - Me $
12	-7	7
14	-5	5
15	-4	4
18	-1	1
19	-	-
20	1	1
21	2	2
24	5	5
25	6	6
Total		31

Calculamos la desviación mediana absoluta:

$$DMe = \frac{31}{9} = 3.44$$

Actividad 4

Del ejemplo 5.3 calcule la desviación mediana absoluta.

Actividad 5

Del ejemplo 5.4 calcule la desviación mediana absoluta.

2.5 La varianza y el desvío estándar

2.5.1 Concepto

Como vimos, el rango se deja influenciar por los valores extremos, y no considera los demás valores de la muestra. Las dos medidas de dispersión, la varianza y el desvío estándar, son las más importantes.

Tanto la varianza como el desvío estándar, son medidas de la variabilidad de las observaciones en relación al centro de la distribución, más precisamente la media aritmética. La variabilidad de cualquier distribución generalmente se contempla en términos de la desviación de cada valor observado (x) con respecto a la media aritmética ($x - \bar{x}$). Si las desviaciones son pequeñas, obviamente los datos son menos

variables o están menos dispersos, que si las desviaciones son grandes. Dado que la sumatoria de los desvíos con respecto a la media es igual a cero, por propiedad de la media, la varianza lo soluciona elevándolos al cuadrado. La fórmula de cálculo de la varianza para una muestra es:

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

NOTA: Si se tratara de una población finita la media aritmética sería sustituida por la media de la población (μ) y el denominador sería N. Pero dado que no es común la media poblacional la reemplazamos por su estimador, es decir, la media aritmética.

Si los datos están agrupados a la sumatoria se la multiplica por las frecuencias absolutas respectivas a cada valor de la variable. Si los datos están agrupados en intervalos de clase se toman los puntos medios o marcas de clase para los desvíos. La fórmula quedaría de esta manera para una muestra:

$$S^2 = \frac{\sum (x_m - \bar{x})^2 f_i}{n-1}$$

El desvío estándar es la raíz cuadrada positiva de la varianza.

$$S = \sqrt{\text{Varianza}}$$

Recordemos que tanto para el cálculo de la varianza como para el desvío estándar, si calculamos para la población, el denominador de las fórmulas es N, es decir el tamaño de la población.

2.5.2 Consideraciones sobre la varianza y el desvío estándar

- 1) La varianza y el desvío estándar miden la dispersión promedio en torno a la media, es decir, cómo fluctúan las observaciones mayores por encima de la media y cómo se distribuyen las observaciones menores por debajo de ella.
- 2) La varianza tiene ciertas propiedades matemáticas útiles. Sin embargo, al calcularla se obtienen unidades al cuadrado: dólares al cuadrado, puntos al cuadrado, etc. Por

ello, en la práctica la principal medida de dispersión que se utiliza es el desvío estándar, cuyo valor está dado en las unidades originales de los datos: dólares, puntos, etc.

- 3) Se elevan al cuadrado a las desviaciones, porque la suma de las desviaciones con respecto a la media siempre es igual a cero. Es decir:

$$x_1 - \bar{x} + x_2 - \bar{x} + \dots + x_{n-1} - \bar{x} + x_n - \bar{x} = 0$$

- 4) Ni la varianza ni el desvío estándar pueden ser negativos, dado que las diferencias se elevan al cuadrado.
- 5) Tanto la varianza como el desvío estándar se calculan sólo cuando se trata de variables cuantitativas.
- 6) Cuanto más separados o dispersos estén los datos, tanto mayores serán el rango, la varianza y el desvío estándar.
- 7) Si los datos están más concentrados, menores serán la varianza y el desvío estándar. En forma inversamente, a mayor valor del coeficiente de la varianza y el desvío estándar, mayor dispersión de los datos respecto a su media aritmética. Por lo tanto, cuando los valores de los coeficientes son menores, el grupo de observaciones de la variable es más “**homogéneo**” que si dichos coeficientes fueran mayores. O sea: “a mayor dispersión menor homogeneidad y a menor dispersión, mayor homogeneidad.
- 8) Si todas las observaciones son iguales, de manera que no haya variación en los datos, el rango, la varianza y el desvío estándar serán iguales a cero.
- 9) Si en un conjunto de valores de una variable, por ejemplo calificaciones, cuanto más elevado sea el desvío estándar, menos representativa será la calificación promedio.
- Al comparar dos distribuciones siempre se enfoca la atención sobre su posición y su variabilidad. Si las posiciones son las mismas, ellas pueden diferir en su variabilidad y, por el contrario, pueden tener la misma variabilidad pero diferir en su posición. Por ejemplo, dos países pueden tener ingresos medios iguales pero con diferentes desviaciones estándares, lo que dará diferentes modelos de distribución de ingresos. Un país cuyos ingresos tiene una desviación estándar pequeña podría indicar que su población está compuesta por una gran clase media. En cambio, si la distribución de los ingresos presenta una desviación estándar muy grande puede estar indicando que dicho país tiene una gran clase baja o alta y una clase media muy pequeña o nula, nos estaría indicando una mayor desigualdad social de la población.
- 10) Se divide por “ $n - 1$ ” porque de esa forma el estimador tiene mejores propiedades que si se dividiera por “ n ”. Ahora al dividir por “ $n - 1$ ” no incide de la misma forma para muestras pequeñas o para muestras grandes, por lo tanto, para muestras grandes $n \geq 60$, puede usarse cualquiera de las dos formas (n ó $n - 1$).

2.5.3 Cálculo de la varianza y el desvío estándar

2.5.3.1 Para datos no agrupados

Para datos no agrupados, utilizamos la expresión que dimos para la definición de la varianza y el desvío estándar.

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1} \rightarrow \text{Varianza}$$

$$S^2 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \rightarrow \text{Desvío Estándar}$$

Ejemplo 5.6

Los pesos de 6 niños de 1 año son:

9, 10, 10, 11, 11 y 12 Kg.

Calcular la varianza y el desvío estándar de los pesos de los niños de la muestra.

Solución:

Se procede a través de los siguientes pasos:

- I. Se calcula la media aritmética de la muestra.
- II. Se hace la diferencia entre cada observación y la media aritmética.
- III. Se elevan al cuadrado las diferencias.
- IV. Se suman los cuadrados de las desviaciones.
- V. La suma se divide por “n – 1” al tratarse de una muestra pequeña.

Los pasos mencionados se resumen en el siguiente cuadro:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
9	-1.5	144
10	-0.5	121
10	-0.5	121
11	0.5	121
11	0.5	121
12	1.5	144
$\sum x_i = 63$	0	$\sum (x_i - \bar{x})^2 = 772$

$$\bar{x} = 63 / 6 = 10.5 \text{ kg}$$

$$S^2 = \frac{772}{5} = 154.4 \text{ kg}^2 \quad \text{Varianza}$$

$$S = \sqrt{154.4} = 12.43 \text{ kg} \quad \text{Desvío estándar}$$

Interpretación: La variabilidad o dispersión de los pesos de los niños con respecto al peso promedio es de 12.43 Kg. Es decir la dispersión de cómo varían los pesos mayores de los niños y los pesos menores que el peso promedio de los niños es de 12.43 Kg.

Ejemplo 5.7

Se analizó en un hospital las frecuencias cardiacas de 8 niños:

130,130, 131, 128, 132, 127, 129, 132 pulsaciones por minuto.

Determinar la varianza y el desvío estándar de la frecuencia cardiaca de la muestra.

Solución:

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
127	-2.875	8.266
128	-1.875	3.516
129	-0.875	0.766
130	0.125	0.016
130	0.125	0.016
131	1.125	1.266
132	2.125	4.516
132	2.125	4.516
$\sum x_i = 1039$	0	22.878

La media es: $1039/8 = 129.875$, es decir 130 pulsaciones por minuto.

Nota: Es importante destacar que el valor de la media tiene tres decimales, pero a los fines de interpretación se redondea. Ahora para realizar los cálculos de la varianza o desvío estándar se utiliza los decimales, tomamos los tres por fines operativos, pero tengamos en cuenta que cuantos más decimales utilicemos obtendremos mayor exactitud. En el ejemplo, la suma de los desvíos no nos hubiera dado "0" si hubiéramos utilizado 130.

La varianza es:

$$S^2 = \frac{22.878}{7} = 3.268 \text{ pulsaciones}^2 \text{ por min.}$$

$$S = \sqrt{3.268} = 1.808 \text{ pulsaciones por min.}$$

El desvío estándar es 1.808 pulsaciones por minuto.

Interpretación del desvío estándar: Las desviaciones de las frecuencias cardiacas de los niños, con respecto a la frecuencia cardiaca promedio, es de 1.808 pulsaciones por minuto.

NOTA

No interpretamos a la varianza, dado que las unidades están elevadas al cuadrado, por lo que pierde sentido.

Actividad 6

Las edades de los pacientes de un pabellón de un hospital son:

38, 27, 14, 41 y 22 años

- Calcular la varianza y el desvío estándar. Interpretar el desvío estándar.
- Si se seleccionara otra muestra en otro pabellón con los siguientes resultados:

17, 28, 32, 45 y 62 años

¿En cuál de las muestras hay mayor dispersión?

2.5.3.2 Para datos agrupados sin intervalos de clase

Como en este caso tenemos la presencia de las frecuencias, debemos considerarlas.

$$S^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{n-1} \rightarrow \text{Varianza}$$

$$S^2 = \sqrt{\frac{\sum (x_i - \bar{x})^2 f_i}{n-1}} \rightarrow \text{Desvío Estándar}$$

Los pasos que debemos seguir para determinar el cálculo de la varianza y del desvío estándar son:

- Se coloca en una columna los valores que toma la variable y en otra las respectivas frecuencias absolutas correspondientes a cada valor.
- Se encuentra la media aritmética de los datos ponderados por sus frecuencias.
- Se halla la diferencia entre cada valor y la media aritmética.
- Se eleva al cuadrado las desviaciones y se multiplica por cada frecuencia. Los valores obtenidos se suman.
- A la suma de los cuadrados de las desviaciones se divide por $n - 1$. Obtenemos la varianza.
- Al cociente anterior se calcula la raíz cuadrada y obtenemos el desvío estándar.

Veamos para su mejor comprensión a través de un ejemplo:

Ejemplo 5.8

La permanencia de pacientes en una clínica en días durante un mes, se registra en la siguiente tabla:



Nº de días	Nº de pacientes
1	15
2	20
3	34
4	26
5	28
6	29
7	50

Calcular el desvío estándar e interpretarlo.

Solución:

Nº de días	Nº de pacientes	$x_i f_i$	F_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
1	15	15	15	-4.27	18.23	273.45
2	20	40	35	-3.27	10.69	213.80
3	34	102	69	-2.27	5.15	175.10
4	26	104	95	-1.27	1.61	41.86
5	28	140	123	-0.27	0.07	1.96
6	29	174	152	0.73	0.53	15.37
7	50	350	202	1.73	2.99	149.50
Total	202	1065		*	39.27	871.04

* Si consideramos todos los decimales nos daría 0.

$$\bar{x} = \frac{1065}{202} = 5.27 \cong 5 \text{ días}$$

$$S^2 = \frac{871.04}{201} = 4.33 \text{ días}^2$$

$$S = \sqrt{4.33} = 2.08 \text{ días}$$

Interpretación del desvío estándar: La desviación o variabilidad de los días de permanencia de los pacientes de la clínica con respecto al número de días promedio es de 2 días.

Actividad 7

En una ciudad se determinó a través de una muestra, los precios de los alquileres de departamentos de dos ambientes y el número de inquilinos que contrataron. Esta información se observa en la siguiente tabla:

Precios de los alquileres \$	Número de inquilinos
300	600
350	720
360	680
400	540
470	476
550	320
600	190

Determinar el desvío estándar e interpretarlo.

2.5.3.3 Para datos agrupados en intervalos de clase

En este caso lo único que varía con respecto a la fórmula anterior, es que los x_i son los puntos medios de cada intervalo de clase.

$$S^2 = \frac{\sum (x_m - \bar{x})^2 f_i}{n-1} \rightarrow \text{Varianza}$$

$$S^2 = \sqrt{\frac{\sum (x_m - \bar{x})^2 f_i}{n-1}} \rightarrow \text{Desvío Estándar}$$

NOTA: Para la varianza, se divide por $n-1$ para que sea el mejor estimador

Ejemplo 5.9

El siguiente cuadro muestra la distribución de acuerdo a su peso expresado en kilogramos, de 60 alumnos:

Peso (Kg.)	20-25	25-30	30-35	35-40	40-45	45-50	50-55
Nº de alumnos	4	10	9	18	7	6	6

Hallar el desvío estándar e interpretarlo.

Solución:

Los pasos que debemos seguir para el cálculo del desvío estándar son:

7. Se determina el punto medio de cada clase.
8. Se encuentra la media aritmética de los datos tabulados.
9. Se halla la diferencia entre cada marca de clase y la media aritmética.
10. Se eleva al cuadrado las desviaciones y se suman.
11. A la suma de los cuadrados de las desviaciones se divide por $n - 1$. Obtenemos la varianza.
12. Al cociente mencionado se calcula la raíz cuadrada y obtenemos el desvío estándar.

Pesos	Xm	fi	fi Xm	Xm - X	(Xm - X) ²	(Xm - X) ² fi
20-25	22.5	4	90.0	-14.67	215.21	860.84
25-30	27.5	10	275.0	-9.67	93.51	935.10
30-35	32.5	9	292.5	-4.67	21.81	196.29
35-40	37.5	18	675.0	0.33	0.11	1.98
40-45	42.5	7	297.5	5.33	28.41	198.87
45-50	47.5	6	285.0	10.33	106.71	640.26
50-55	52.5	6	315.0	15.33	235.01	1410.06
Totales		60	2230			4243.40

La media aritmética es: $2230/60 = 37.17$ Kg.

La varianza es: $4243.40/59 = 71.92$ Kg. al cuadrado

El desvío estándar es: $\sqrt{71.92} = 8.48$ Kg.

Interpretación: La variabilidad de los pesos de los escolares, con respecto al peso promedio es de 8.48 Kg.

Actividad 8

Los siguientes datos corresponden al alquiler mensual pagado por inquilinos de 211 departamentos:



Alquileres (\$)	Número de inquilinos
350 - 380	3
380 - 410	8
410 - 440	15
440 - 470	13

470 - 500	33
500 - 530	48
530 - 560	35
560 - 590	28
590 - 620	16
620 - 650	12

Calcule el desvío estándar e interprételo.

2.5.3.4 Propiedades de la varianza

I. Propiedad 1: La varianza de un conjunto de datos $x_1, x_2, \dots, x_n$ siempre es un número no negativo. Esto es:

$$S^2(x) \geq 0$$

Comprobación:

$$S^2(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Como $(x_i - \bar{x})^2 \geq 0$ para todo i , y además $n-1 > 0$, nos queda:

$$S^2(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \geq 0.$$

II. Propiedad 2: La varianza de una constante es cero.

Si $x_1 = x_2 = \dots = x_n = a$, donde a es una constante, tenemos:

$$S^2(a) = 0$$

Comprobación:

Si $x_i = a$ para $i = 1, 2, 3, \dots, n$. Entonces:

$$\bar{a} = a \quad y,$$

$$(x_i - \bar{a})^2 = (a - \bar{a})^2 = (a - a)^2 = 0$$

Y si dividimos por $n-1$, nos queda:

$$S^2(a) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{a})^2 = 0$$

III. Propiedad 3: Si a cada observación $x_1, x_2, \dots, x_n$ se suma o se resta una constante $a \geq 0$, la varianza del nuevo conjunto de valores coincide con la varianza del conjunto original.

$$S^2(y) = S^2(x \pm a) = S^2(x)$$

Comprobación:

Vamos a designar a “y” el nuevo valor resultante de la suma o resta de la constante a cada valor de x, es decir:

$$y_i = x_i + a, \text{ por lo tanto}$$

$$\bar{y} = \bar{x} + a$$

Ahora la varianza nos queda:

$$S^2(y) = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i + a - \bar{x} - a)^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = S^2(x).$$

IV. Propiedad 4: Si a cada valor de un conjunto se multiplica por una constante a , la varianza del nuevo conjunto de valores, es igual a la varianza del conjunto original multiplicada por el cuadrado de la constante.

$$\underline{\underline{S^2(y) = a^2 S^2(x)}}$$

Comprobación:

Vamos a designar a “y” el nuevo valor resultante del producto de la constante a cada valor de x, es decir:

$$y_i = ax_i, \text{ entonces } \bar{y} = a\bar{x}$$

Luego:

$$\begin{aligned} S^2(y) &= \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n-1} \sum_{i=1}^n (ax_i - a\bar{x})^2 = \frac{1}{n-1} \sum_{i=1}^n a^2 (x_i - \bar{x})^2 \\ &= \frac{a^2}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = a^2 S^2 \end{aligned}$$

2.5.3.5 Método abreviado para el cálculo de la varianza

Este método tiene como objetivo disminuir la cantidad de operaciones en el cálculo de la varianza para datos agrupados en clase. Operando algebraicamente sobre la definición de la varianza, utilizando las propiedades de la sumatoria y la definición de la media se llega a una fórmula de trabajo de la varianza:

$$S^2 = \frac{\sum_{i=1}^n x_i^2 f_i - \frac{\left(\sum_{i=1}^n x_i f_i\right)^2}{n}}{n-1}$$

O lo que es lo mismo si tenemos calculada la media:

$$S^2 = \frac{\sum_{i=1}^n x_i^2 f_i - \bar{x}^2 n}{n-1}$$

Ejemplo 5.10

Del ejemplo 5.9 utilizar la fórmula de trabajo de la varianza.

Peso (Kg.)	20-25	25-30	30-35	35-40	40-45	45-50	50-55
Nº de alumnos	4	10	9	18	7	6	6

$$S^2 = \frac{\sum_{i=1}^n x_i^2 f_i - \bar{x}^2 n}{n-1}$$

Pesos	Xm	x ²	fi	x ² fi	fi Xm
20-25	22.5	506.25	4	2025.00	90.0
25-30	27.5	756.25	10	7562.50	275.0
30-35	32.5	1056.25	9	9506.25	292.5
35-40	37.5	1406.25	18	25312.50	675.0
40-45	42.5	1806.25	7	12643.75	297.5
45-50	47.5	2256.25	6	13537.50	285.0
50-55	52.5	2756.25	6	16537.50	315.0
Totales			60	87125.00	2230.00

La media aritmética es: $2230/60 = 37.17$

La varianza es:

$$S^2 = \frac{87125 - 82896.6}{59} = 71.67$$

Las diferencias se deben al redondeo en los decimales. Si se utilizaran todos los decimales posibles, ambos resultados serían idénticos.

😊 **Importante:** Las propiedades de la varianza son análogas del desvío estándar.

2.5.3.6 Teorema de Tchebyshev

Dado un conjunto de observaciones $x_1, x_2, \dots, x_n$, entre la media aritmética y $\pm k$ veces el desvío estándar existe por lo menos, el

**$100(1 - 1/k^2) \%$ con $k \geq 1$
de las observaciones sin que importe, como se distribuye el conjunto de datos.**

Lo que nos dice el teorema es que en el intervalo $\bar{x} - kS$ y $\bar{x} + kS$ se encuentra por lo menos el $100(1 - 1/k^2) \%$ de las observaciones o datos.

Para $k = 1$, no se puede precisar.

Para $k = 2$. Entre $\bar{x} - 2S$ y $\bar{x} + 2S$ deben estar, cuando menos $100(1 - 1/2^2) = 75\%$ de las observaciones o $3/4$ de las mismas.

Para $k = 3$: dentro de $\bar{x} - 3S$ y $\bar{x} + 3S$ deben estar, cuando menos $100(1 - 1/3^2) = 88.89 \%$ de las observaciones.

Para $k = 4$. Entre $\bar{x} - 4S$ y $\bar{x} + 4S$ estarán contenidas, cuando menos $100(1 - 1/4^2) = 93.75 \%$ de las observaciones.

En el teorema de Tchebyshev se menciona “cuando **menos**”, esto se debe a que este teorema es aplicable a cualquier distribución. En la práctica se da con gran frecuencia que la fracción de observaciones que caen en el intervalo especificado supera a $1 - 1/k^2$. Es por ello la utilización de una regla que describe con exactitud llamada “**Regla empírica**”, aplicable a una distribución de datos cuya forma es acampanada y simétrica u otras formas parecidas a ella, que en la mayoría de los casos en la práctica su ocurrencia es muy frecuente, de ahí la utilización de esta regla. La distribución a que se hace mención, que se verá en otro capítulo posterior, se llama distribución normal. Veamos que nos dice la regla empírica:

Dada una distribución de un conjunto de observaciones de tamaño n suficientemente grande, que es simétrica y forma aproximadamente acampanada se verifica lo siguiente:

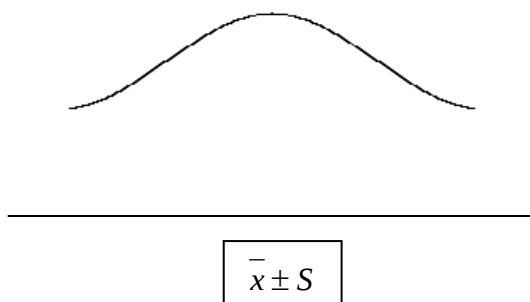
Entre $\bar{x} - S$ y $\bar{x} + S$ se encuentra el 68.27 % de las observaciones.

Entre $\bar{x} - 2S$ y $\bar{x} + 2S$ se encuentra el 95.45 % de las observaciones.

Entre $\bar{x} - 3S$ y $\bar{x} + 3S$ se encuentra el 99.73 % de las observaciones.

Entre $\bar{x} - 4S$ y $\bar{x} + 4S$ se encuentra el 99.99 % de las observaciones.

Esta regla práctica o empírica tiene la utilidad de describir la dispersión o variabilidad de un gran número de datos. La distribución de los datos en forma acampanada, llamada distribución normal adopta la siguiente forma:



Ejemplo 5.11

La dirección de un colegio solicitó que se determine el promedio de los alumnos en la materia de Matemáticas (500 alumnos). Los resultados fueron que el promedio fue de 7.12 puntos y el desvío estándar de 2.33 puntos. Suponemos que la distribución de los datos es acampanada. Aplique la regla empírica y el teorema de Tchebyshev y compare.

Solución:

$$\bar{x} - S = 7.12 - 0.33 = 6.79$$

$$\bar{x} + S = 7.12 + 0.33 = 7.45$$

$$\bar{x} - 2S = 7.12 - 0.66 = 6.46$$

$$\bar{x} + 2S = 7.12 + 0.66 = 7.78$$

$$\bar{x} - 3S = 7.12 - 0.99 = 6.13$$

$$\bar{x} + 3S = 7.12 + 0.99 = 8.11$$

Según la regla práctica, entre 6.79 y 7.45 se encuentra el 68.27% de las calificaciones en Matemáticas. Entre 6.46 y 7.78 se encuentra el 95.45% de las calificaciones; y entre 6.13 y 8.11 se encuentra el 99.73% de las notas en Matemáticas de los alumnos del colegio.

Aplicando el teorema de Tchebyshev tenemos que entre 6.46 y 7.78 se encuentran cuando menos el 75% de los datos. Entre 6.13 y 8.11 se encuentran cuando menos el 88.89% de los datos.

Observaciones: Podemos observar que si aplicamos el teorema de Tchebyshev los porcentajes de los datos son menores, esto se debe a que dicho teorema contempla cualquier distribución. Si la distribución fuera acampanada o aproximada, la regla práctica se ajustaría mejor. También es importante destacar que si el desvío estándar fuera mayor los intervalos tendrían una mayor amplitud es decir los extremos estarían más distanciados e implica menor precisión.

Actividad 9

¿Si el estudio realizado en el colegio del ejemplo 5.11 también se realiza en otro colegio y se obtiene el mismo promedio pero con un desvío estándar igual a 2.33, que conclusión podría sacar?,

2.6 El coeficiente de variación

2.6.1 Concepto y cálculo

El coeficiente de variación es una medida relativa de dispersión. Es el cociente entre el desvío estándar y la media aritmética, multiplicado por 100.

$$CV = \frac{S}{X} \times 100$$

Nota

Al multiplicar por 100 se convierte la expresión decimal en porcentaje.

Dado que surge como el cociente entre dos cantidades expresadas en la misma unidad de medida, es independiente de ella. El coeficiente de variación es útil sobre todo cuando se compara la variabilidad. Si queremos comparar dos o más medidas de dispersión (por ejemplo, el desvío estándar de los ingresos de los empleados de una fábrica y el desvío estándar de las inasistencias para el mismo grupo), resulta imposible. No podemos comparar pesos y días de inasistencia al trabajo. Para poder comparar, necesitamos convertir cada una de estas medidas a un valor relativo que no dependa de las unidades, es decir, a un porcentaje. Esta medida es el **coeficiente de variación (CV)**.

Ejemplo 5.12

Un estudio de las calificaciones obtenidas en un curso interno sobre principios de administración, y los años de servicio de los empleados que tomaron el curso, dio como resultado los siguientes valores estadísticos: la calificación promedio fue de 200 puntos; el desvío estándar, 40 puntos. El promedio del número de años de servicio fue

de 20 años, y el desvío estándar de 2 años. Comparar la dispersión relativa de las dos distribuciones:

$$\begin{aligned} CV_1 &= 40/200 \text{ (100)} \\ &= 20\% \end{aligned}$$

$$\begin{aligned} CV_2 &= 2/20 \text{ (100)} \\ &= 10\% \end{aligned}$$

Interpretación: De acuerdo a los resultados obtenidos, podemos observar que se evidencia mayor dispersión en las calificaciones que en los años de antigüedad.

Ejemplo 5.13

El director de un colegio desea determinar la variabilidad de las calificaciones entre dos divisiones de primer grado. Se realizan los estudios y dan los siguientes resultados:

Primer Grado "A": $CV = 5.4$

Primer Grado "B": $CV = 9.6$

¿En cuál división hay mayor variabilidad?

Con este ejemplo, podemos inferir que en el Primer Grado "B", hay mayor dispersión en las calificaciones, con respecto al "A". Además podemos decir que la calificación promedio en el "A" es más representativa. Podríamos haber determinado los desvíos estándar respectivos, y hubiéramos llegado a la misma conclusión para este caso en particular.

Actividad 10

En una empresa se tomó una prueba de inteligencia a sus empleados y dio un puntaje promedio de 88 puntos y un desvío estándar de 12 puntos. También se encontró que el número promedio de inasistencias por mes es de 2 días, y un desvío estándar de 0.43 días. Compare la variabilidad relativa para informar al Director de Recursos Humanos

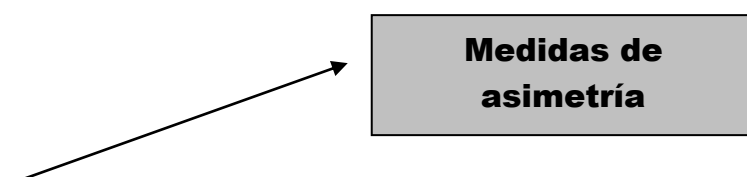
3. Asimetría y Curtosis

Medidas de forma:

- Asimetría
- Curtosis o apuntamiento.

Hasta ahora, hemos estado analizando y estudiando la dispersión de una distribución, pero parece evidente que necesitamos conocer más sobre la forma de una distribución. En esta parte, analizaremos las medidas de forma, a través de un histograma o representación de datos y qué información nos aporta según la forma que tengan las observaciones.

Las medidas de forma de una distribución se pueden clasificar en dos grandes grupos o bloques: **medidas de asimetría y medidas de curtosis**.



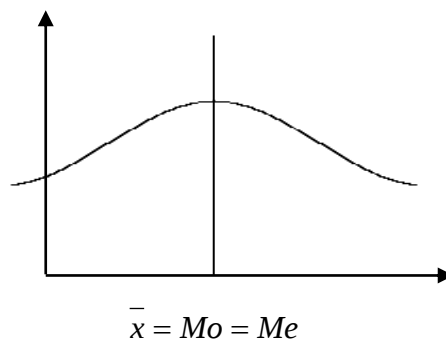
MEDIDAS DE FORMA

**Medidas de
curtosis**
3.1 Medidas de asimetría y curtosis

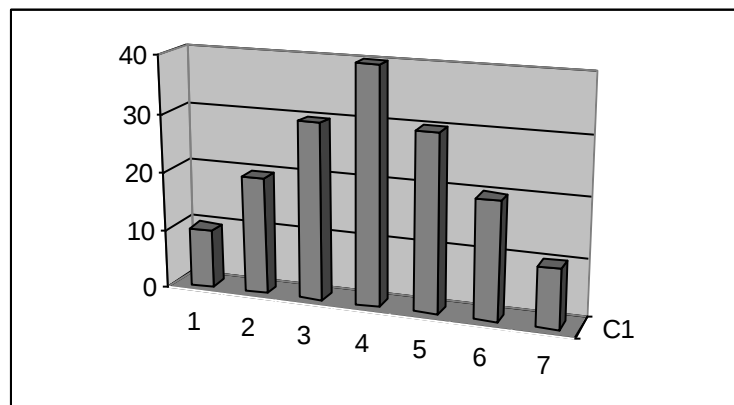
Cuando al trazar una vertical, en el polígono de frecuencias unimodales o en el histograma de una variable, por el valor de la media, esta vertical, se transforma en eje de simetría. Diremos pues, que es simétrica, cuando a ambos lados de la media aritmética haya el mismo número de valores de la variable, equidistantes de dicha media, y tales que dichos valores equidistantes tiene la misma frecuencia absoluta.

En caso contrario, dicha distribución será asimétrica o diremos que presenta asimetría. Cuando la distribución es simétrica, la media, la moda y la mediana coinciden.

$$\bar{x} = Mo = Me$$



También la simetría podemos observarla en un gráfico de barras:



Existen dos tipos de asimetría:

- Asimetría positiva

➤ Asimetría negativa

Asimetría positiva

Los datos en una distribución se encuentran concentrados a la izquierda y dispersos hacia la derecha. La media aritmética es mayor que la moda, y la mediana queda comprendida entre ambos.

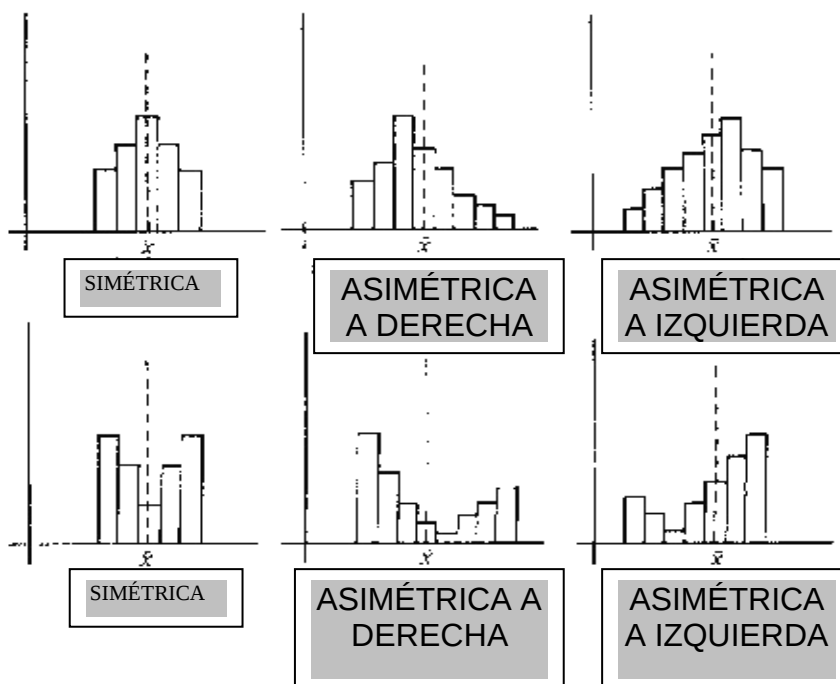
$$Mo < Me < \bar{x}$$

Asimetría negativa

En una distribución de datos, los mismos se encuentran concentrados a la derecha y dispersos hacia la izquierda. La media es menor que la moda y la mediana, ésta última sigue permaneciendo en el centro de ambas medidas. La mediana en ambos casos se encuentra en el centro, al dividir los datos en dos partes iguales.

$$\bar{x} < Me < Mo$$

Histogramas



Para calcular la asimetría, una posibilidad, es utilizar el llamado **coeficiente de FISHER** que representaremos como **g₁** y responderá a la siguiente expresión matemática:

$$g_1 = \frac{\sum (x_i - \bar{x})^3 f_i}{ns^3}$$

Según sea el valor de g_1 , diremos que la distribución es asimétrica positiva, negativa, o simétrica, o sea:

Si $g_1 > 0 \rightarrow$ la distribución será asimétrica positiva o sesgada hacia la derecha (desplazada hacia la derecha).

Si $g_1 < 0 \rightarrow$ la distribución será asimétrica negativa o sesgada hacia la izquierda (desplazada hacia la izquierda).

Si $g_1 = 0 \rightarrow$ la distribución será simétrica.

Otra forma de calcular la asimetría es utilizando el **coeficiente de PEARSON**, el cual simbolizamos como **As**:

$$As = \frac{\bar{x} - Mo}{S}$$

Interpretación:

Si $As > 0 \rightarrow$ la distribución será asimétrica positiva o sesgada hacia la derecha (desplazada hacia la derecha).

Si $As < 0 \rightarrow$ la distribución será asimétrica negativa o sesgada hacia la izquierda (desplazada hacia la izquierda).

Si $As = 0 \rightarrow$ la distribución será simétrica.

Ejemplo 5.14

Del ejemplo 5.9 calcular los coeficientes de asimetría, interpretarlos y compararlos.

El siguiente cuadro muestra la distribución de acuerdo a su peso expresado en kilogramos, de 60 alumnos:

Peso (Kg.)	20-25	25-30	30-35	35-40	40-45	45-50	50-55
Nº de alumnos	4	10	9	18	7	6	6

Solución:

Pesos	X_m	f_i	$f_i X_m$	$x_m - \bar{x}$		$(X_m - \bar{x})^2 f_i$
20-25	22.5	4	90.0	-14.67	215.21	860.84
25-30	27.5	10	275.0	-9.67	93.51	935.10
30-35	32.5	9	292.5	-4.67	21.81	196.29
35-40	37.5	18	675.0	0.33	0.11	1.98
40-45	42.5	7	297.5	5.33	28.41	198.87
45-50	47.5	6	285.0	10.33	106.71	640.26
50-55	52.5	6	315.0	15.33	235.01	1410.06

Totales		60	2230			4243.40
----------------	--	-----------	-------------	--	--	----------------

La media aritmética es: $2230/60 = 37.17$ Kg.

La varianza es: $4243.40/59 = 71.92$ Kg. al cuadrado

El desvío estándar es: $\sqrt{71.92} = 8.48$ Kg.

Curtosis

El **Coefficiente de Curtosis** analiza el grado de concentración que presentan los valores que toma la variable alrededor de la zona central de la distribución.

Se definen 3 tipos de distribuciones según su grado de curtosis:

- 1. Distribución mesocúrtica:** presenta un grado de concentración medio alrededor de los valores centrales de la variable (casi el mismo que presenta una distribución normal).
- 2. Distribución leptocúrtica:** presenta un elevado grado de concentración alrededor de los valores centrales de la variable.
- 3. Distribución platicúrtica:** presenta un reducido grado de concentración alrededor de los valores centrales de la variable.

Leptocúrtica

Mesocúrtica

Platicúrtica

La fórmula de cálculo para medir la curtosis es la siguiente:

$$k = \frac{0,5 (Q_3 - Q_1)}{P_{90} P_{10}}$$

Si el resultado de la fórmula tiende o se aproxima a 0.5 la curva es leptocúrtica, si tiende a 0.25 es mesocúrtica, y si tiende a 0 la curva es platicúrtica.

4.0 Indicadores de control

Actividad 1

Los siguientes datos corresponden a cuentas por cobrar (en miles de \$) de una empresa. Calcular el rango.

22 31 31.5 38 40 41 50.8 72

Resolución;

$$R = 72 - 22 = 50$$

Actividad 2

Se selecciona una muestra de 11 obreros de una fábrica, y se les pregunta las horas extras que realizaron en la semana:

2 4 4 5 5.5 6 6 6 6.5 3 3

- ✓ Calcular el rango inter-cuartílico.
- ✓ ¿Qué ocurre si uno de ellos se equivocó y se rectifica afirmando que hizo 12 horas ? ¿ El resultado o conclusión es igual ?

Resolución;

✓

Ordenamos de menor a mayor los datos:

2 3 3 4 4 5 5.5 6 6 6 6.5

$N = 11$

$$\frac{n+1}{4} = \frac{12}{4} = 3$$

$$\frac{3(n+1)}{4} = \frac{36}{4} = 9$$

$$Q_1 = 3$$

$$Q_3 = 6$$

$$RI = Q_3 - Q_1 = 6 - 3 = 3$$

✓

El resultado es el mismo ya que no considera en el cálculo los valores extremos, y sólo el 50% central de la distribución. De todas maneras aumenta la dispersión, pero no es captada por esta medida.

Actividad 3

Las edades de los suscriptores de una revista especializada en cine, se observan en el siguiente cuadro. Calcular la desviación media absoluta.

Edades	Nº de suscriptores	Marca de clase	$x_i f_i$	$x_m - \bar{x}$	$ x_m - \bar{x} f_i$
15-20	23	17.5	402.5	-24.28	558.44
20-25	47	22.5	1057.5	-19.28	906.16
25-30	92	27.5	2530.0	-14.28	1313.76
30-35	105	32.5	3412.5	-9.28	974.40
35-40	250	37.5	9375.0	-4.28	1070.0
40-45	240	42.5	10200.0	0.72	172.80
45-50	190	47.5	9025.0	5.72	1086.80
50-55	156	52.5	8190.0	10.72	1672.32
55-60	120	57.5	6900.0	15.72	1886.40
Total	1223		51092.5		9641.08

$$\bar{x} = \frac{51092.5}{1223} = 41.78 \cong 42 \text{ años}$$

$$DM = \frac{9641.08}{1223} = 7.88 \text{ años}$$

Actividad 4

Del ejemplo 5.3 calcule la desviación mediana absoluta.

En una localidad de la provincia se aprobaron ocho proyectos del programa social "Componentes materiales Tipología VI". El número de beneficiarios del Plan Jefes/as de Hogar por cada proyecto se especifica en orden ascendente:

1 2 2 3 3 3 4 4

Resolución:

La mediana es igual a 3.

Número de beneficiarios	Desviación $x_i - Me$	Desviación absoluta $ x_i - Me $
1	-2	2
2	-1	1
2	-1	1
3	-	-
3	-	-
3	-	-
4	1	1
4	1	1
Total		6

$$DM = \frac{6}{8} = 0.75$$

Actividad 5

Del ejemplo 5.4 calcule la desviación mediana absoluta.

Los datos que se ven en el siguiente cuadro agrupados en intervalos de clase corresponden a la cantidad de préstamos otorgados a pequeñas y medianas empresas, en un período determinado, por entidades financieras de una ciudad, el objetivo de los mismos fue la adquisición de bienes de capital y capital de trabajo. Calcule la dispersión media absoluta:

Nº de préstamos	Nº de empresas	Marca de clase	$F_i \succ$	Desvíos $x_m - Me$	Desviación absoluta $ x_m - Me $	$ x_m - Me f_i$
1 – 5	17	3	17	-5	5	85
5 – 9	19	7	36	1	1	19
9 – 13	13	11	49	3	3	39
13 – 17	9	15	58	7	7	63
17 – 21	6	19	64	11	11	66
21 – 25	2	23	66	15	15	30
N	66					302

$$Me = 5 + \left(\frac{33 - 17}{19} \right) \times 4 =$$

$$= 8.37 \cong 8 \text{ préstamos}$$

$$DM = \frac{302}{66} = 4.58$$

Actividad 6

Las edades de los pacientes de un pabellón de un hospital son:

38, 27, 14, 41 y 22 años

- Calcular la varianza y el desvío estándar. Interpretar el desvío estándar.
- Si se seleccionara otra muestra en otro pabellón con los siguientes resultados:

17, 28, 32, 45 y 62 años

¿En cuál de las muestras hay mayor dispersión?

Resolución:

a)

La media aritmética es:

$$\bar{x} = \frac{142}{5} = 28.4 \cong 28 \text{ años}$$

Edad de los pacientes	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
14	-14.4	207.36
22	-6.4	40.96
27	-1.4	1.96
38	9.6	92.16
41	12.6	158.76
Total	0	501.2

$$S^2 = \frac{501.2}{4} = 125.3 \text{ años}^2$$

$$S = \sqrt{125.3} = 11.19 \text{ años}$$

Interpretación: La variabilidad o dispersión de las edades de los pacientes con respecto a la edad promedio es de 11 años.

17, 28, 32, 45 y 62 años

b)

La media aritmética es:

$$\bar{x} = \frac{184}{5} = 36.8 \cong 37 \text{ años}$$

Edad de los pacientes	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
17	-19.8	392.04
28	-8.8	77.44
32	-4.8	23.04
45	8.2	67.24
62	25.2	635.04
Total	0	1194.8

$$S^2 = \frac{1194.8}{5} = 238.96 \text{ años}^2$$

$$S = \sqrt{238.96} = 15.46 \text{ años}$$

En la segunda muestra es mayor la dispersión al ser mayor el desvío estándar, lo que significa mayor variabilidad en las edades de los pacientes. Nos permite comparar.

Actividad 7

En una ciudad se determinó a través de una muestra, los precios de los alquileres de departamentos de dos ambientes y el número de inquilinos que contrataron. Esta información se observa en la siguiente tabla:

Precios de los alquileres \$	Número de inquilinos
300	600
350	720
360	680
400	540
470	476
550	320
600	190

Determinar el desvío estándar e interpretarlo

Resolución:

Precios de los alquileres \$	Número de inquilinos	$x_i f_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 f_i$
300	600	180000	-99.01	9802.98	5881788.0
350	720	252000	-49.01	2401.98	1729425.60
360	680	244800	-39.01	1521.78	1034810.40
400	540	216000	0.99	0.98	529.20
470	476	223720	70.99	5039.58	2398840.08
550	320	176000	150.99	22797.98	7295353.60
600	190	114000	200.99	40396.98	7675426.20
Total	3526	1406520			26016173.08

$$\bar{x} = \frac{1406520}{3526} = 399.01$$

$$S^2 = \frac{26016173.08}{3525} = 7380.47$$

$$S = \sqrt{7380.47} = 85.91$$

La dispersión o variabilidad de los precios de los alquileres con respecto al precio promedio del alquiler de los departamentos es de \$ 85.91.

Actividad 8

Los siguientes datos corresponden al alquiler mensual pagado por inquilinos de 211 departamentos:

Alquileres (\$)	Número de inquilinos
350 - 380	3
380 - 410	8
410 - 440	15
440 - 470	13
470 - 500	33
500 - 530	48
530 - 560	35
560 - 590	28
590 - 620	16
620 - 650	12

Calcule el desvío estándar e interprételo.

Resolución:

350 - 380	3
380 - 410	8
410 - 440	15
440 - 470	13
470 - 500	33
500 - 530	48
530 - 560	35
560 - 590	28
590 - 620	16
620 - 650	12

Precio de los alquileres \$	Nº de Inquilinos f_i	x_m	$x_m f_i$	$x_m - \bar{x}$	$(x_m - \bar{x})^2$	$(x_m - \bar{x})^2 f_i$
350-380	3	365	1095	-124.79	15572.54	46717.62
380-410	8	395	3160	-94.79	8985.14	71881.12
410-440	15	425	6375	-64.79	4197.74	62966.10
440-470	13	455	5915	-34.79	1210.34	15734.42
470-500	33	485	16005	-4.79	22.94	757.02
500-530	48	515	24720	25.21	635.54	30505.92
530-560	35	545	19075	55.21	3048.14	106684.9
560-590	28	575	16100	85.21	7260.74	203300.72
590-620	16	605	3280	115.21	13273.34	212373.44
620-650	12	635	7620	145.21	21085.94	253031.28
Total	211		103345			1003952.54

$$\bar{x} = \frac{103345}{211} = 489.79 \text{ pesos}$$

$$S^2 = \frac{1003952.54}{210} = 4780.73$$

$$S = \sqrt{4780.73} = 69.14 \text{ pesos}$$

La variabilidad del precio de los alquileres mensuales con respecto al alquiler mensual promedio es de 69.14 pesos.

Actividad 9

¿Si el estudio realizado en el colegio del ejemplo 5.11 también se realiza en otro colegio y se obtiene el mismo promedio pero con un desvío estándar igual a 2.33, que conclusión podría sacar?,

La dirección de un colegio solicitó que se determine el promedio de los alumnos en la materia de Matemáticas (500 alumnos). Los resultados fueron que el promedio fue de 7.12 puntos y el desvío estándar de 2.33 puntos. Suponemos que la distribución de los datos es acampanada. Aplique la regla empírica y el teorema de Tchebyshev y compare.

Resolución:

$$\bar{x} - S = 7.12 - 2.33 = 4.79$$

$$\bar{x} + S = 7.12 + 2.33 = 9.45$$

$$\bar{x} - 2S = 7.12 - 4.66 = 2.46$$

$$\bar{x} + 2S = 7.12 + 4.66 = 11.78$$

$$\bar{x} - 3S = 7.12 - 6.99 = 0.13$$

$$\bar{x} + 3S = 7.12 + 6.99 = 14.11$$

Según la regla práctica, entre 4.79 y 9.45 se encuentra el 68.27% de las calificaciones en Matemáticas. Entre 2.46 y 11.78 se encuentra el 95.45% de las calificaciones; y entre 0.13 y 14.11 se encuentra el 99.73% de las notas en Matemáticas de los alumnos del colegio.

Aplicando el teorema de Tchebyshev tenemos que entre 2.46 y 11.78 se encuentran cuando menos el 75% de los datos. Entre 0.13 y 14.11 se encuentran cuando menos el 88.89% de los datos.

Podemos observar que si aplicamos el teorema de Tchebyshev los porcentajes de los datos son menores, esto se debe a que dicho teorema contempla cualquier distribución. Si la distribución fuera acampanada o aproximada, la regla práctica se ajustaría mejor. También es importante destacar que si el desvío estándar fuera mayor los intervalos tendrían una mayor amplitud, es decir los extremos estarían más distanciados e implica menor precisión.

Actividad 10

En una empresa se tomó una prueba de inteligencia a sus empleados y dio un puntaje promedio de 88 puntos y un desvío estándar de 12 puntos. También se encontró que el número promedio de inasistencias por mes es de 2 días, y un desvío estándar de 0.43 días. Compare la variabilidad para informar al Director de Recursos Humanos

Resolución:

$$CV_1 = \frac{12}{88} = 0.14 \times 100 = 14\%$$

$$CV_2 = \frac{0.43}{12} = 0.036 \times 100 = 3.6\%$$

Hay menor variabilidad en el número de inasistencias por mes. Se puede afirmar que en los dos casos no hay tanta variabilidad, no están muy alejados del 0%, sobre todo en el segundo caso. De acuerdo a lo afirmado estamos en presencia de promedios representativos.

8.0 Ejercicios propuestos

1- Una fábrica instala dispositivos de seguridad en portones y puertas. El tiempo de instalación en minutos llevado a cabo en 9 portones fue:

16 – 18 – 19.5 – 20 – 23 – 24.5 – 25 – 28 – 30

- Calcular el rango y el rango inter-cuartílico.
- Calcular la desviación media y la desviación mediana. Comparar.

2- Una empresa tiene que enviar 10 cajas de encomienda a una ciudad. Los pesos de las cajas, en kilogramos, son: 30, 32, 36, 44, 52, 68, 71, 103, 110 y 124.

- Calcular la varianza y el desvío estándar e interpretarlos.
- Si en otro envío se registraran los mismos valores, salvo el mayor, en vez de ser 124 es 190. ¿La variabilidad de los pesos se ve afectada? Compruébelo.

3 – El número de inasistencias de los alumnos de un colegio durante un mes, se observa en el siguiente cuadro:

Nº de inasistencias	Nº de alumnos
1	52
2	60
3	80
4	43
5	36
6	24
7	18

Calcule la variabilidad de los datos.

4- El ingreso promedio de una ciudad surgido de una muestra es de \$480.- y un desvío estándar de \$23.-. Aplique el teorema de Tchebyshev y la regla empírica. Si el desvío estándar fuera \$ 49.-, comente.

5- Una empresa consideró la conveniencia de implantar 2 programas de capacitación para realizar idéntica tarea para sus empleados y se seleccionaron dos grupos.

En el grupo en que se implementó el programa 1 se requirieron en promedio 22,37 horas para capacitar a cada empleado, con una varianza de 57,19. En el grupo en donde se llevó a cabo el programa 2 se necesitó un promedio de 18,05 horas para capacitar a cada empleado, con una varianza de 69,02. ¿Cuál programa presentó la menor variabilidad relativa en sus resultados?

6- Los sueldos de una gran empresa sin considerar niveles jerárquicos se detallan a continuación:

Sueldos \$	Nº de empleados
400 – 500	320
500 – 600	380
600 – 700	425
700 – 800	629
800 – 900	370
900 – 1000	126
1000 – 1100	94

Realice un diagnóstico sobre la variabilidad de los sueldos con las medidas de dispersión vistas.

PARA MEDIDAS DE CURTOSIS O APUNTAMIENTO

6.- ¿Cuándo aplico muestreo de juicio? R.- Cuando un individuo elige los elementos a incluir en la muestra. Ejemplo: En lugar de elegir aleatoriamente las personas a encuestar, se van eligiendo según se vea dicha persona, en físico, en carácter, etc.

¿Cuándo aplico muestreo por conveniencia? R.- cuando se buscan resultados óptimos sin arriesgarse a obtener malos, por lo que se eligen elementos de la muestra que nos aseguren resultados positivos. Ejemplo: Si quiero saber la preferencia de tomar coca cola, pregunto a puras personas que vea comprando o tomando este refresco.

¿Cuándo aplico muestreo por subgrupos racionales? R.- Se aplica en procesos donde tomando una muestra pequeña en intervalos de tiempo advertimos cambios en la calidad. Ejemplo: Si
M en I Juan Jose Hurtado Moreno
2
un proceso donde se producen 100 botellas de agua, quiero saber que tantos defectos tienen, muestreo de
20 en 20 en intervalos de tiempos para así ver los cambios o las mejoras entre cada muestra.

¿Cuándo aplico muestreo sistemático? R.- Se aplica cuando queremos un resultado en orden, es decir, se seleccionan elementos de la población a un intervalo uniforme en una lista ordenada. Ejemplo: para detectar la iluminación pública de un vecindario donde cada duodécima casa ocupa un esquina.

¿Cuándo aplico muestreo estratificado? R.- Cuando los elementos de una población se clasifican primero en grupos, por características. Ejemplo: Si se quisiera formar equipos parejos en actitudes y capacidades, tendríamos que primero clasificar a los estudiantes de un grupo en subgrupos con características similares y luego aplicar muestreo simple.

Distribución mesocúrtica: presenta un grado de concentración medio alrededor de los valores centrales de la variable (el mismo que presenta una distribución normal).

Distribución leptocúrtica : presenta un elevado grado de concentracion alrededor de los valores centrales de la variable.

Distribución platicúrtica: presenta un reducido grado de concentracion alrededor de los valores centrales de la variable.