FEPOL Fondo Editorial

biblioteca nacional del perú

https://editorialfondo.com/

Medidas de Dispersión

Florencio Flores Ccanto
Lourdes Gálvez Morales
César Lau Mego
Marianella Marilú Villegas Lira
Vicente Carlos Dávila Huamán
Isabel Menacho Vargas

Professionals On line S.A.C.

MEDIDAS DE DISPERSIÓN

Florencio, Flores Ccanto
Lourdes, Gálvez Morales
César, Lau Mego
Marianella Marilú, Villegas Lira
Vicente Carlos, Dávila Huamán
Isabel, Menacho Vargas

Lima – Perú

MEDIDAS DE DISPERSIÓN

© **Florencio Flores Ccanto**
Dirección: Cooperativa de Vivienda Pablo Patrón Mz: Ñ, Lote 13, Lurigancho, Lima - Perú
fflores@une.edu.pe

Lourdes Gálvez Morales
Dirección: Mz: J3, Lote: 9. A.H. Yanacoto Parcela 3-5ta Zona, Lurigancho, Lima - Perú
lgalvez@une.edu.pe

César Lau Mego
Dirección: Rua Loefgreen, 1057 - 4° Andar Vila Clementino, Sao Paulo, Brasil
cesar.slmg@gmail.com

Marianella Marilú Villegas Lira
Dirección: Urb. Santa Raquel, Mz:C, Lote: 16, Ate Vitarte, Lima - Perú
mvillegas@une.edu.pe

Vicente Carlos Dávila Huamán
Dirección: Jr. Restauración 388, Lima – Perú
vdavila@une.edu.pe

Isabel Menacho Vargas
Dirección: Rio Tambo 185, Pueblo Libre, Lima – Perú
imenachov@unmsm.edu.pe

Editada por:
© Professionals On Line SAC. (FEPOL) - Fondo Editorial.
Dirección: Av. La Marina Nro: 2900, San Miguel - Perú
professionalsonline.net@gmail.com
Teléf. móvil: +51 999 140 920
Web: https://professionalsonline.net/

Coeditor
Biblioteca Nacional del Perú
Dirección: Av. De La Poesía 160, 15034 San Borja - Lima, Perú

Primera edición digital: Noviembre 2022
Libro digital disponible en: https://editorialfondo.com/

Hecho el Depósito Legal en la Biblioteca Nacional del Perú N° 2022-10979
ISBN: 978-612-48981-7-4
DOI: https://doi.org/10.47422/fepol.8

Corrección de estilo: Luis Pablo Diaz Tito
luisp.diaz@upsjb.edu.pe / Tel. de contacto: +51 955 129 801

Diseño y Diagramación: Gráfica “imagen”
Manuel Enrique Sampen Antonio
sampen25@gmail.com / Tel. de contacto: +51 990 064 589

Libro resultado de Investigación y con revisión por pares doble ciego.
Sello editorial: Fondo Editorial (978-612-48981)

COTENIDO

AGRADECIMIENTOS

Agradecimiento a todos los integrantes del grupo de investigación “Didácticos Matemáticos e Informáticos” de la Universidad Nacional de Educación Enrique Guzmán y Valle.

PRÓLOGO

El libro presenta didácticamente los contenidos de medidas de dispersión de datos, en marco del proceso enseñanza-aprendizaje de la estadística básica, mediante el uso del emulador de la calculadora científica Casio fx-570LA-X Classwiz, que facilita las operaciones a la medida de dispersión, varianza, desviación estándar, coeficiente de variación, coeficiente de asimetría para datos simples y agrupados.

El Capítulo I, desarrolla los conceptos y aplicaciones del estudio de medidas de dispersión, rango, rango intercuartil y desviaciones; cada tema se presenta paso a paso, acompañado de ejemplos para datos simples y agrupados.

El Capítulo II, presenta los conceptos de la varianza, propiedades de la varianza, cálculo de la varianza para datos simples y agrupados, utilizando ejemplos contextualizados.

El Capítulo III, presenta los conceptos desviación estándar, ejemplos aplicados a situaciones didácticas de la vida cotidiana.

En el Capítulo IV, expone los conceptos coeficiente de variación y sus aplicaciones a situaciones didácticas que les servirán en el desarrollo de la investigación científica.

El Capítulo V, presenta el estudio del coeficiente de asimetría y sus aplicaciones en la solución de situaciones didácticas aplicados a la investigación y finalmente presenta ejemplos y ejercicios resueltos.

Utilidad del libro

Este libro es útil para comprender los conceptos, propiedades del estudio de dispersión y el estudio de la variabilidad de los datos y su interpretación de los mismos. Asimismo, para resolver los conflictos semióticos del proceso enseñanza-aprendizaje en los estudiantes.

Daniel M. Chirinos M.
Vicerrector de Investigación de la UNE

RESUMEN

El estudiante se familiariza con los contenidos que le permita comprender los temas de medidas de dispersión, varianza, desviación estándar, coeficiente de variación y coeficiente de asimetría, haciendo uso como herramienta informática el emulador Casio fx-570LA-X Classwiz en el proceso de enseñanza-aprendizaje que el docente debe utilizar para desarrollar los contenidos antes mencionados. Una de las bondades del emulador es que hace muy didáctico abordar cada tema haciéndola más sencilla, amena, divertida y les crea mucho interés con los temas de estadística. Los ejemplos y ejercicios desarrollados en el presente trabajo son situaciones reales y casos motivadores. El libro termina presentando ejercicios adicionales para que el estudiante los pueda abordar en sus tiempos libres.

Palabras clave: Dispersión, varianza, desviación estándar, coeficiente de variación y coeficiente de asimetría.

CAPÍTULO I

Medida de Dispersión

Definición. - Mide la cantidad típica en que los valores del conjunto de datos difieren de la medida aritmética (Hanke y Reitsh, 1996) "consisten en números que otorgan información acerca de la variabilidad de los datos” (Vílchez, p. 111)

MEDIDAS DE DISPERSIÓN

Mide la dispersión de los datos respecto de la media aritmética (Hanke y Reitsh, 1996).

VARIANZA (S^2)

DESVIACIÓN ESTÁNDAR (S)

Para datos sin agrupar:

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

Para datos agrupados:

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

Para datos sin agrupar:

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

Para datos agrupados:

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

Definición. - Las medidas de dispersión o variabilidad son números que muestran el grado de separación de los datos respecto a un valor central, que generalmente es la media aritmética. (Mayorga-Ponce, Reyes-Torres., Baltazar-Téllez & Martínez-Alamilla, 2021).

Para medir el grado de dispersión de una variable, se utilizan generalmente los siguientes indicadores: el rango, la varianza, la desviación estándar o desviación típica y el coeficiente de variación.

Si se tienen dos grupos de datos con igual media aritmética, no significa que estas sean iguales, debemos analizar el grado de homogeneidad entre sus datos. Así, por ejemplo, los datos 10, 40 y 70 tienen igual media aritmética y mediana que los datos 39, 40 y 41; sin embargo, en el primer grupo la media se encuentra alejada de los valores extremos 10 y 70, cosa que no ocurre con el segundo grupo. Se observa, entonces, que el primer grupo de datos es más heterogéneo o dispersa que la segunda.

Para medir el grado de dispersión de una variable, se utilizan generalmente los siguientes indicadores: el rango, la varianza, la desviación estándar o desviación típica y el coeficiente de variación.

1.2 El rango

Definición. - El rango o recorrido se define como la diferencia entre los datos mayor y menor de una distribución. En forma de ecuación, lo expresamos por:

Rango = Dato mayor – Dato menor (Pagano, p.76)

Notación. Se denota con $R =$ rango, X_{max} = dato mayor y X_{min} = dato menor; por tanto,

$$R = X_{max} - X_{min}$$

donde X_{max} y X_{min} son los valores máximo y mínimo de la variable X, respectivamente.

Tomando los datos del ejemplo anterior, se tiene:

$$R_1 = 70 - 10 = 60 \quad \text{y} \quad R_2 = 41 - 40 = 1$$

Se observa la gran dispersión de la primera distribución respecto de la homogeneidad de la segunda.

Ejemplo 1. Si para un conjunto de datos se tiene que $X_{max} =$ 300 y $X_{min} = 100$, entonces se tiene

Que: $R = 300 - 100 = 200$.

Notas:

a) El rango es una medida de dispersión fácil de calcular.

b) El rango es una medida de dispersión muy inestable, ya que su valor cambia mucho ante datos extremos y depende únicamente de estos valores extremos. Esto es, el rango mide únicamente la dispersión de los datos extremos y no la de los datos intermedios.

c) Se calcula para variables medidas en escala de intervalo o de razón.

Ejemplo 2. Calcular el rango de los siguientes grupos de datos:

a) 4; 8; 11; 15; 18; 20 $\quad$ R = 20 – 4 = 16

b) 36; 22; 17; 45; 28; 37; 19 $\quad$ R = 45 – 17 = 28

c) 2,6; 5,1; 3,7; 7,7; 4,1 $\quad$ $R = 7{,}7 - 2{,}6 = 5{,}1$

Rango Intercuartil (RI)

Definición. El rango intercuartil (RI) es la diferencia entre el tercer cuartil (Q_3) y el primer cuartil (Q_1). Esto es,

$$RI = Q_3 - Q_1 \qquad \text{(Molina \& Rodrigo, 2010, p.6)}$$

Ejemplo 3. En una tabla sobre la distribución de 80 trabajadores según sus sueldos, se encontró que el $Q_1 =$ 1840,80 soles y $Q_3 =$ 2310,20 soles. Hallar el rango intercuartil.

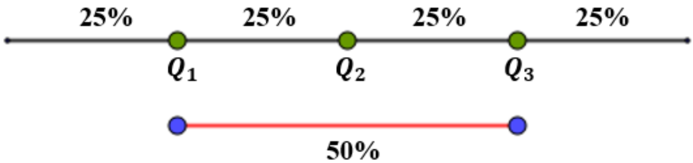
Solución:

$$RI = Q_3 - Q_1 = 2310{,}20 - 1840{,}80 = 469{,}40$$

Este valor significa que el 50 % de los sueldos de los trabajadores, varía en un intervalo de 469,40 soles.

Notas:

1. En el RI se encuentra el 50 % de los datos y no se encuentran afectados por los valores extremos. Es una medida que excluye el 25 % más alto y el 25 % más bajo, como se observa en el gráfico.



2. Si el RI es muy pequeño entonces, describe alta uniformidad o pequeña variabilidad de los valores centrales.

Puntajes de desviación

Definición. - El puntaje de desviación nos indica que tan lejos está el dato original con respecto a la media aritmética de su distribución.

En forma de ecuación, el puntaje de desviación (Pagano, 2004, p.77) se define como:

$X - \bar{X}$: puntaje de desviación para datos de una muestra, donde $\bar{x}$ es la media muestral.

$X - \mu$: puntaje de desviación para los datos de una población, donde μ es la media poblacional.

Ejemplo 4. Se tienen los datos: 2, 4, 6, 8, 10. La media aritmética de estos datos es 6.

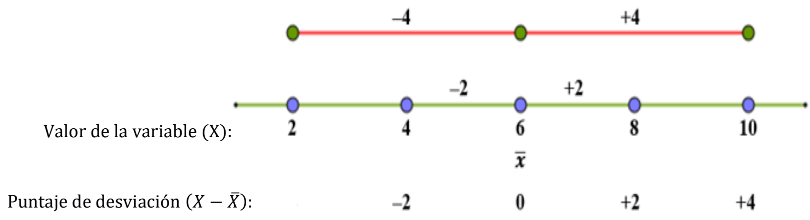
$$\bar{X} = \frac{\sum x_i}{n} = \frac{30}{5} = 6$$

En la siguiente tabla, tenemos el cálculo de los puntajes de desviación

$\boldsymbol{x_i}$	$\boldsymbol{x_i - \overline{X}}$	Puntajes de desviación
2	2 – 6	– 4
4	4 – 6	– 2
6	6 – 6	0
8	8 – 6	2
10	10 – 6	4
$\sum x_i = 30$		Suma de los puntajes de desviación $= 0$

Se observa en la tabla que la suma de las desviaciones es igual a cero.

En la gráfica se muestran los datos originales y sus correspondientes puntajes de desviación.



CAPÍTULO II

Varianza

La varianza

La varianza cuantifica el grado de dispersión o de variación de los valores de una variable cuantitativa respecto de su media aritmética. Es una medida de variabilidad cuyo valor nos indicará si los datos están bastantes concentrados o dispersos con respecto a su media aritmética.

Definición. - La varianza se define como la media aritmética de los cuadrados de las desviaciones de los datos con respecto a su media.

Notas:

1. La varianza calculada a partir de una muestra se denota con S^2 y referido a una población se denota con σ^2 (sigma minúscula al cuadrado), O también,

   $V(X)$: varianza de la variable X. (sea muestral o poblacional)

   S^2: varianza muestral ordinaria.

2. Cuanto mayor sea la magnitud de la varianza los datos estarán más dispersos o diseminados con respecto a la media (más heterogéneos), y cuanto menor sea la varianza los datos estarán más concentrados alrededor de la media (más homogéneos).

Notación. Cuando se calcula la varianza de una población estadística se simboliza por σ^2 (sigma cuadrada). Cuando se refiere a datos de una muestra se denota por S^2. En la calculadora Casio fx-570LA-X Classwiz la notación es $\sigma^2 x$ para la varianza poblacional y $S^2 x$ para la varianza muestral.

Cálculo de la varianza

La varianza es una medida de la dispersión que se calcula hallando la media aritmética o promedio de los cuadrados de las diferencias de los datos con respecto a su media aritmética (desviaciones). Entonces:

$$\sigma^2 = \frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + (x_3 - \bar{X})^2 + \cdots + (x_n - \bar{X})^2}{n} = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n}$$

Esto es, $\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n}$

Varianza de datos no agrupados

Sean los n datos originales de una muestra de una variable cuantitativa X: $x_1, x_2, x_3, \dots, x_n$ cuya media es $\bar{X}$ y desviaciones $(x_i - \bar{X})$ con $i = 1, 2, \cdots, n$. Entonces, la varianza de estos n valores es el número positivo

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

Se desarrolla la fórmula anterior:

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

Obtenemos una equivalente que nos permite hallar la varianza con mayor comodidad. Esta fórmula se expresa por

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

Ejemplo 5. Con los siguientes datos: 11, 10, 14, 16, 12, 18 y 9, calcular la varianza.

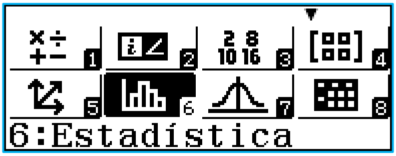
Solución:

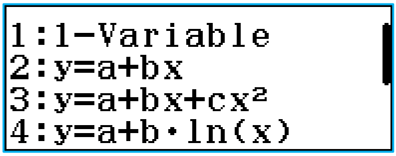
Para calcular la varianza emplearemos la información presentada en la siguiente tabla, y nos basaremos en su expresión (fórmula).

$\boldsymbol{i}$	$\boldsymbol{x_i}$	$\boldsymbol{x_i^2}$
1	11	121
2	10	100
3	14	196
4	16	256
5	12	144
6	18	324
7	9	81
Total	$\sum_{i=1}^{7} x_i = 90$	$\sum_{i=1}^{7} x_i^2 = 1222$

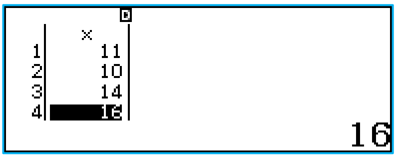
- $\bar{X} = \frac{\sum_{i=1}^{7} x_i}{7} = 12{,}857$
- $\frac{\sum_{i=1}^{7} x_i^2}{7} = 174{,}57$
- $\sigma^2 = \frac{\sum_{i=1}^{7} x_i^2}{7} - \bar{x}^2 = 9{,}27$

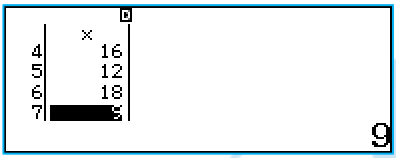
En la calculadora Classwiz se pueden ingresar los datos en el modo Estadística:



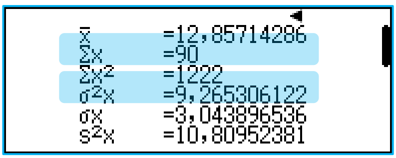


Seleccionamos la opción 1-Variable e ingresamos los datos:





T3



La calculadora presenta en pantalla la suma de los cuadrados de los datos ingresados y la media:

$\sum x^2 = 1222$

$\bar{X} = 12{,}857$

La varianza se puede calcular utilizando estos valores, tal como vimos anteriormente.

$$\sigma^2 = \frac{\sum x^2}{7} - \bar{X}^2$$

Reemplazando los valores y calculando, se obtiene:

$$\frac{1222}{7} - (12{,}857)^2$$

$$9{,}268979571$$

Además, se puede observar que la calculadora presenta el valor de la varianza poblacional

$$\sigma^2 x = 9{,}265 \ldots$$

Puede observar los resultados obtenidos con su emulador o calculadora Classwiz escaneando el siguiente código QR:

Untitled Paper

$\bar{x} = 12.85714286$
$\Sigma x = 90$
$\Sigma x^2 = 1222$
$\sigma^2 x = 9.265306122$
$\sigma x = 3.043896536$
$s^2 x = 10.80952381$
$sx = 3.287784027$
$n = 7$
$\min(x) = 9$
$Q_1 = 10$
$Med = 12$
$Q_3 = 16$
$\max(x) = 18$

Ejemplo 6. Encontrar la varianza de los siguientes datos:

1, 1, 2, 2, 2, 3, 4, 5, 5, 9, 10

Solución:

Primero se debe calcular es la media de los datos; luego, usar la primera o la segunda fórmula.

$$\bar{X} = \frac{\sum x_i}{n} = \frac{44}{11} = 4$$

Luego,

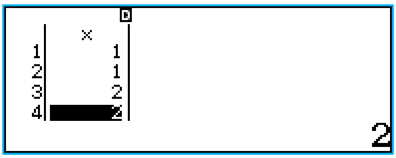
$$\sigma^2 = \frac{\sum_{i=1}^{11} x_i^2}{11} - \bar{x}^2 = \frac{1^2 + 1^2 + 2^2 + 2^2 + 2^2 + 3^2 + 4^2 + 5^2 + 5^2 + 9^2 + 10^2}{11} - 4^2$$

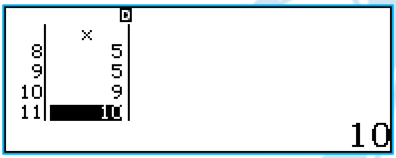
$$\sigma^2 = \frac{1 + 1 + 4 + 4 + 4 + 9 + 16 + 25 + 25 + 81 + 100}{11} - 16$$

$$\sigma^2 = \frac{270}{11} - 16$$

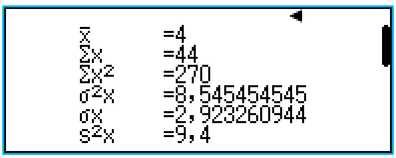
$$\sigma^2 = 8{,}5454$$

Utilizando el emulador de la calculadora Classwiz:





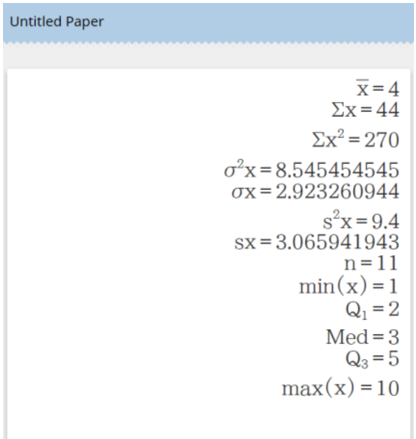
T3



La varianza poblacional es:

$$\sigma^2 x = 8,\widehat{54}$$

Puede observar los resultados obtenidos con su emulador o calculadora Classwiz escaneando el siguiente código QR:


Untitled Paper

$\bar{x} = 4$
$\Sigma x = 44$
$\Sigma x^2 = 270$
$\sigma^2 x = 8.545454545$
$\sigma x = 2.923260944$
$s^2 x = 9.4$
$sx = 3.065941943$
$n = 11$
$\min(x) = 1$
$Q_1 = 2$
$Med = 3$
$Q_3 = 5$
$\max(x) = 10$

Ejemplo 7. Dos grupos de estudiantes, *A* y *B*, rindieron una prueba de razonamiento matemático y sus calificaciones fueron las siguientes:

Grupo A	12	10	13	9	11
Grupo B	13	12	14	10	6

¿Cuál de los grupos presenta menor viariabilidad?

Solución:

Calculamos el promedio de calificaciones en cada grupo:

$$\bar{X}_A = \frac{12 + 10 + 13 + 9 + 11}{5} = 11$$

$$\bar{X}_B = \frac{13 + 12 + 14 + 10 + 6}{5} = 11$$

Como los promedios son iguales, la menor variabilidad la podemos medir con la varianza.

$$\sigma_A^2 = \frac{12^2 + 10^2 + 13^2 + 9^2 + 11^2}{5} - 11^2 = 123 - 121 = 2$$

$$\sigma_B^2 = \frac{13^2 + 12^2 + 14^2 + 10^2 + 6^2}{5} - 11^2 = 129 - 121 = 8$$

Por tanto, el Grupo A presenta menor variabilidad de sus calificaciones.

Varianza de datos agrupados

Si los datos se encuentran agrupados en una tabla de frecuencias la varianza se puede calcular, teniendo en cuenta el tipo de variable.

- **Para una variable discreta**

La varianza de n valores de una variable estadística discreta X, que se clasifican en k valores distintos x_1, x_2, $\cdots$, x_k y cuya media aritmética es $\bar{X}$, se calcula utilizando la fórmula:

$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i\,(x_i - \bar{X})^2}{n}, \text{ donde}$$

σ^2: Varianza.

x_i: Valor de la variable X.

$\bar{X}$: Media aritmética de los datos.

f_i: Frecuencia absoluta del valor x_i.

n: Tamaño de la muestra.

Como en el caso anterior, de esta fórmula se deduce

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

Ejemplo 8. La siguiente tabla indica el número de hijos por familia. Calcular la varianza.

$\boldsymbol{x_i}$	$\boldsymbol{f_i}$
0	5
1	12
2	20
3	8
4	5
Total	n = 50

Solución:

1. Cálculo de la media ($\bar{X}$)

$\boldsymbol{x_i}$	$\boldsymbol{f_i}$	$\boldsymbol{f_i . x_i}$
0	5	0
1	12	12
2	20	40
3	8	24
4	5	20
Total	n = 50	96

$$\bar{X} = \frac{96}{50} = 1,92$$

2. Cálculo de la varianza usando la segunda fórmula. Para ello, se completa la tabla de distribución de frecuencia con los valores que nos indica la fórmula.

x_i	f_i	$f_i . x_i$	$f_i . x_i^2$
0	5	0	0
1	12	12	12
2	20	40	80
3	8	24	72
4	5	20	80
Total	n = 50	96	244

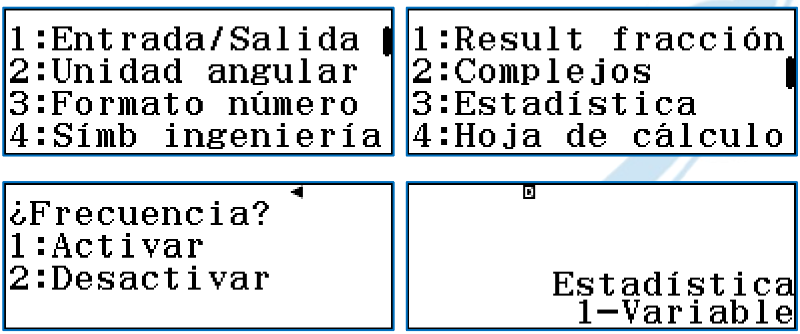
$$\sigma^2 = \frac{\sum_{i=1}^{k} f_i \, x_i^2}{n} - \bar{x}^2 = \frac{244}{50} - (1{,}92)^2$$

$$\sigma^2 = 1{,}1936$$

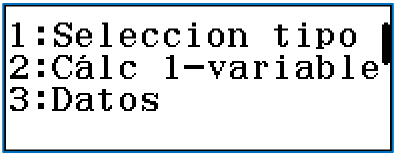
Utilizando el emulador de la calculadora Classwiz:

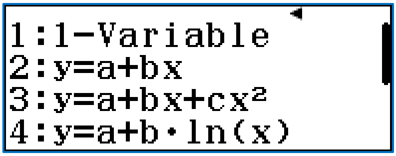
Activación de la columna f (frecuencia)

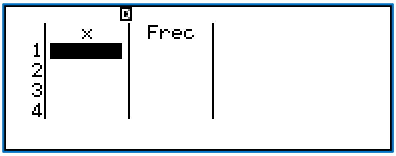
qwR31



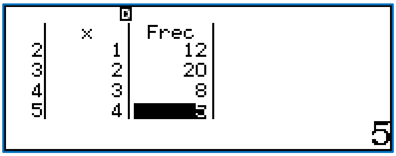
T11



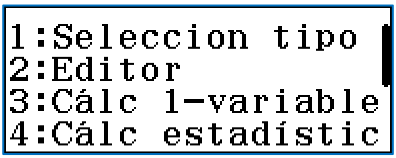


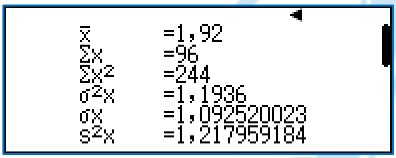


Se cargan los datos en ambas columnas:



T3





Se obtiene el mismo valor para la varianza poblacional:

$$\sigma^2 x = 1{,}1936$$

- **Para una variable continua**

La varianza de n valores de alguna variable X, tabulados en m intervalos de clase; con marcas de clase x_1, $x_2, \cdots, x_m$, frecuencias absolutas $f_1, f_2, \cdots, f_m$ y con media $\bar{x}$ es el número

$$\sigma^2 = \frac{\sum_{i=1}^{m} f_i (x_i - \bar{X})^2}{n} \text{ o } \sigma^2 = \frac{\sum_{i=1}^{m} f_i x_i^2}{n} - \bar{X}^2, \text{ donde}$$

σ^2: Varianza.

x_i: Punto medio de los intervalos de clase.

$\bar{X}$: Media aritmética de los datos.

f_i: Frecuencia absoluta del valor x_i.

n: Tamaño de la muestra.

m: número de intervalos de clase.

Ejemplo 9. Con los datos de la siguiente tabla de distribución de frecuencias (pesos en kg), calcular la varianza.

Clases	$\dot{x}_i$	f_i
[38, 48⟩	43	3
[48, 58⟩	53	2
[58, 68⟩	63	8
[68, 78⟩	73	16
[78, 88⟩	83	9
[88, 98]	93	12
Total		$n = 50$

Solución:

Usando la segunda fórmula, se halla primero la media y luego la varianza. Para ello se agrega dos columnas a la tabla anterior.

Clases	$\dot{x}_i$	f_i	$f_i.x_i$	$f_i.x_i^2$
[38, 48⟩	43	3	129	5547
[48, 58⟩	53	2	106	5618
[58, 68⟩	63	8	504	31752
[68, 78⟩	73	16	1168	85264
[78, 88⟩	83	9	747	62001
[88, 98]	93	12	1116	103788
Total		n = 50	3770	293970

$$\bar{x} = \frac{3770}{50} = 75.4 \text{ kg}$$

$$\sigma^2 = \frac{\sum_{i=1}^{6} f_i\, x_i^2}{n} - \bar{x}^2 = \frac{293970}{50} - (75,4)^2$$

$$\sigma^2 = 194,24 \text{ kg}^2.$$

Nota. La varianza se calcula, también, con frecuencias relativas (o porcentajes). Pues, siendo $h_i = \frac{f_i}{n}$ en la fórmula de la varianza se tiene

$$\sigma^2 = \sum_{i=1}^{m} h_i\, x_i^2 - \bar{x}^2$$

donde $\bar{X}^2 = \sum_{i=1}^{m} h_i\, x_i$.

Ejemplo 10. Con los datos de la tabla anterior, calcular la varianza usando frecuencias relativas.

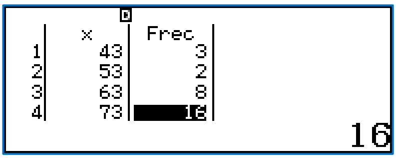
Solución:

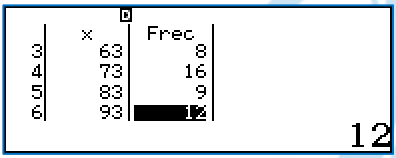
En la tabla de distribución de frecuencias, se calcula las frecuencias relativas.

Clases	$\dot{x}_i$	f_i	h_i	$h_i.\dot{x}_i$	$h_i.\dot{x}_i^{\,2}$
[38, 48⟩	43	3	0,06	2,58	110,94
[48, 58⟩	53	2	0,04	2,12	112,36
[58, 68⟩	63	8	0,16	10,08	635,04
[68, 78⟩	73	16	0,32	23,36	1705,28
[78, 88⟩	83	9	0,18	14,94	1240,02
[88, 98]	93	12	0,24	22,32	2075,76
Total		n = 50	1,00	75,40	5879,40

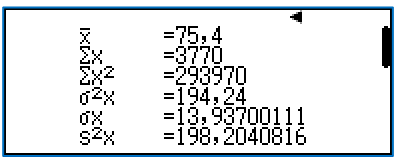
- Media: $\bar{X} = \sum_{i=1}^{6} h_i x_i = 75{,}4\ kg$
- Varianza: $S^2 = \sum_{i=1}^{m} h_{i.}\, x_i^2 - \bar{x}^2 = 5879{,}4 - (75{,}4)^2 \Rightarrow S^2 = 194{,}24\ kg^2$

Uso del emulador de la calculadora Classwiz: Para determinar la varianza poblacional basta ingresar como datos las marcas de clase y sus respectivas frecuencias.





Obteniendo como resultado:



$$\sigma^2 x = 194{,}24$$

Descripción de algunas propiedades de la varianza

i) La varianza siempre es un número no negativo.

ii) Es sensible a valores extremos, grandes o pequeños.

iii) Al sumar a los valores de una de una variable una constante, el valor de la varianza no se altera.

Por ejemplo, sean los valores de una variable X: 3, 5, 7, 9, 11. Entonces, $\bar{X} = 7$ y

$$\sigma_x^2 = \frac{3^2 + 5^2 + 7^2 + 9^2 + 11^2}{5} - 7^2 = \frac{9 + 25 + 49 + 81 + 121}{5} - 49 = 57 - 49 = 8$$

Si sumamos la constante 2 a los valores de la variable X, obtenemos los valores de una variable Y: 5, 7, 9, 11, 13. Entonces, $\bar{y} = 9$ y

$$\sigma_y^2 = \frac{5^2 + 7^2 + 9^2 + 11^2 + 13^2}{5} - 9^2 = \frac{25 + 49 + 81 + 121 + 169}{5} - 81 = 89 - 81 = 8$$

iv) Al multiplicar los valores de una variable por una constante, la varianza queda multiplicada por el cuadrado de dicha constante.

A pesar de su importancia y de la utilidad de la varianza esta presenta dos grandes desventajas:

1. Es un número grande en relación con el valor de los datos, por lo que casi siempre resulta difícil trabajar con ella.
2. Como los resultados vienen dados en unidades elevadas al cuadrado, no tendría ningún sentido su interpretación.

Estas desventajas se pueden obviar si se trabaja con la desviación estándar o desviación típica.

Estas desventajas se pueden obviar si se trabaja con la desviación estándar o desviación típica.

CAPÍTULO III

Desviación Estándar

Como medidas de dispersión la desviación estándar (o la varianza) muestran su utilidad cuando se emplean de manera comparativa. Conocer el valor de la varianza de una distribución no nos permite interpretar de manera inmediata si esta es más o menos dispersa, ya que su valor depende de la unidad de medida de la variable.

Junto con la media aritmética, la desviación estándar es el parámetro más significativo de una población estadística. Entre sus múltiples usos tenemos la aplicación del teorema de Chebyshev y la descripción de la distribución normal, temas que se verán más adelante.

La desviación estándar o desviación típica, denotada por S, es una medida de dispersión que se calcula hallando la raíz cuadrada de la varianza. Esto es,

$$\sigma = \sqrt{\frac{(x_1 - \bar{X})^2 + (x_2 - \bar{X})^2 + (x_3 - \bar{X})^2 + \cdots + (x_n - \bar{X})^2}{n}}$$

Definición. - La desviación estándar, llamada también desviación típica, es la raíz cuadrada positiva de la varianza. Es decir,

$$\sigma = \sqrt{\sigma^2}$$

Notación. La desviación estándar calculada a partir de una muestra se denotará por S y referida a una población por σ, esto es

- $S = \sqrt{S^2}$: desviación estándar de una muestra
- $\sigma = \sqrt{\sigma^2}$: desviación estándar de una población.

Observación. La varianza es una medida de dispersión que tiene unidades al cuadrado, por definición. Por ejemplo, de una varianza de 16 m^2 extrayendo la raíz cuadrada se obtiene una medida de dispersión en las unidades originales; esto es, una desviación estándar de 4 m. Por esta razón se considera a la desviación estándar una medida más usual que la varianza.

Ejemplo 11. Con los valores de una variable X: 3, 5, 7, 9, 11, el valor de la varianza es $S^2 = 8$; por tanto, la desviación estándar será

$$\sigma = \sqrt{\sigma^2} = \sqrt{8} = 2{,}8284$$

Ejemplo 12. Calcule la desviación estándar de los siguientes datos. ¿Por qué σ es tan grande en la parte b, en comparación con la parte a?

a. 6, 8, 7, 3, 6, 4

b. 6, 8, 7, 3, 6, 35

Solución:

Tanto en *a* como en *b*, calculamos primero la varianza y luego la desviación estándar.

a. Media: $\bar{X} = 5{,}67$.

Varianza:

$$\sigma^2 = \frac{6^2 + 8^2 + 7^2 + 3^2 + 6^2 + 4^2}{6} - (5{,}67)^2 = 35 - 32{,}1489 = 2{,}8511$$

Desviación estándar: $\sigma = \sqrt{S^2} = \sqrt{2{,}8511} = 1{,}6885$

b. Media: $\bar{X} = 10{,}83$.

Varianza:

$$\sigma^2 = \frac{6^2 + 8^2 + 7^2 + 3^2 + 6^2 + 35^2}{6} - (10{,}83)^2 = 236{,}5 - 117{,}2889 = 119{,}21$$

Desviación estándar: $\sigma = \sqrt{\sigma^2} = \sqrt{119{,}21} = 10{,}9183$

Respuesta: σ es grande en la parte *b*, porque la desviación estándar es sensible a los valores extremos y 35 es un valor extremo.

Ejemplo 13. Las edades de los miembros de una familia son: 5, 10, 12, 32 y 34 años. Hallar el rango, la varianza y la desviación estándar.

Solución:

- Para hallar el rango se calcula la diferencia entre el mayor y el menor dato. Esto es,

$$R = 34 - 5 = 29$$

Cálculo de la varianza.

$$\bar{X} = \frac{5+10+12+32+34}{5}$$

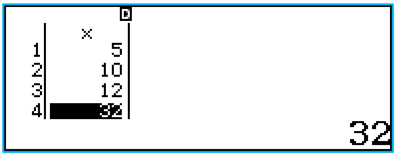
$$\bar{X} = 18{,}6 \text{ años}$$

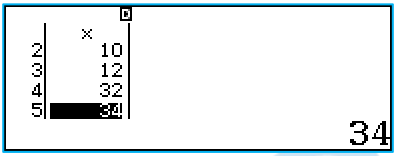
$$\sigma^2 = \frac{5^2+10+12^2+32^2+34^2}{5} - (18{,}6)^2 = 489{,}8 - 345{,}96 = 143{,}84 \text{ años}^2$$

Cálculo de la desviación estándar.

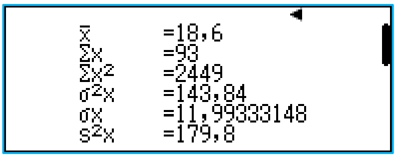
$$\sigma = \sqrt{\sigma^2} = \sqrt{143{,}84} \Rightarrow \sigma = 11{,}99 \text{ años.}$$

Comprobemos los valores obtenidos usando el emulador de la calculadora Classwiz:





T3



$$\sigma x = 11{,}99$$

Puede ver los resultados de cálculo escaneando el siguiente código QR

Ejemplo 14. El costo de producción X de una muestra de cierto tipo de objeto tiene una desviación estándar de \$ 30. El costo medio de producción es de \$ 250 para el 60 % de la muestra y de \$ 200 para el resto. Si su precio de venta en dólares es dado por la relación

$$y = 1{,}1\, x + 10$$

Calcule la media y la varianza del precio de venta de la muestra.

Solución:

Encontramos el promedio de costo de producción usando frecuencia relativa.

$$\bar{X} = 0{,}6(250) + 0{,}4(200) = 150 + 80 = 230$$

Luego, $\bar{X} = 230$ \$.

Media y varianza del precio de venta:

Media: $\bar{y} = 1{,}1\,\bar{x} + 10 = 1{,}1(230) + 10 = 253 + 10 = 263$

Luego, $\bar{Y} = 263$ \$.

Varianza: $S_x = \sqrt{S_x^2} \Rightarrow 30 = \sqrt{S_x^2} \Rightarrow 900 = S_x^2$

Luego,

$y = 1{,}1\, x + 10 \Rightarrow S_y^2 = (1{,}1)^2\, S_x^2$

$= 1{,}21\,(900)$

$= 1\,089$

Por tanto, $S_x^2 = 1089$ \$.

CAPÍTULO III

Coeficiente de Variación

Es una medida de dispersión relativa libre de unidades de medida por lo que es útil para comparar la variabilidad de dos o más grupos de datos expresados en distintas unidades de medida o cuando los promedios de los conjuntos de datos a comparar son diferentes.

Definición. - El coeficiente de variación, denotado por CV, es una medida de dispersión relativa que se define como el cociente entre la desviación estándar y la media aritmética de los datos de una muestra o una población. Esto es,

$$CV = \frac{S}{\bar{X}} \quad \text{o} \quad CV = \frac{\sigma}{\mu}$$

donde,

- $CV = \frac{S}{\bar{X}}$, es el coeficiente de variación de una muestra.
- $CV = \frac{\sigma}{\mu}$, es el coeficiente de variación de una población.

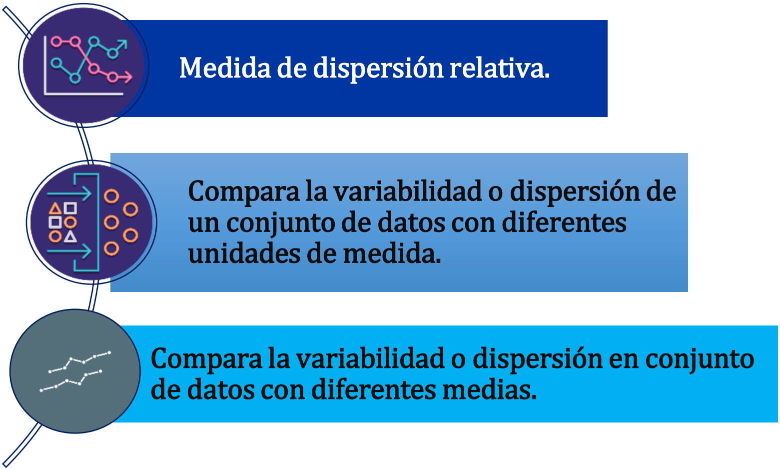
Nota. El coeficiente de variación es un número libre de unidades de medida, motivo por el cual se puede expresar en porcentajes. Es decir,

$$CV = \frac{S}{\bar{X}}(100)\%$$

El coeficiente de variación resulta muy útil para comparar la variabilidad de dos o más conjuntos de datos que tengan diferentes unidades de medida y/o diferentes medias aritméticas. Por ejemplo, si la desviación estándar de las calificaciones obtenidas en la asignatura de Estadística aplicada, en cada una de las secciones C1 y C2 ha sido 3, no se puede afirmar que ambas secciones tienen la misma variabilidad, ya que pueden tener medias diferentes. Si la media de la sección C1 es 12 y la media de la sección C2 es 14, los coeficientes de variación respectivos son:

- $CV_{C1} = \frac{S}{\bar{X}}.(100) = \frac{3}{12}.(100) = 25\ \%$
- $CV_{C2} = \frac{S}{\bar{X}}.(100) = \frac{3}{14}.(100) = 21{,}43\ \%$

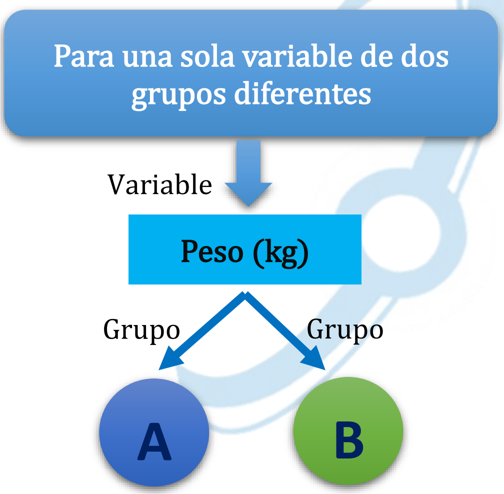
De acuerdo a los resultados, las calificaciones obtenidas en la sección C2 son más homogéneas o tienen menor variabilidad que las calificaciones obtenidas en la sección C1. Según Batanero y Godino (2005) el coeficiente de variación es una medida de dispersión relativa e independiente de la unidad utilizada en los valores de la variable y se pueden comparar distribuciones cuyos datos estén medidos en distintas unidades.



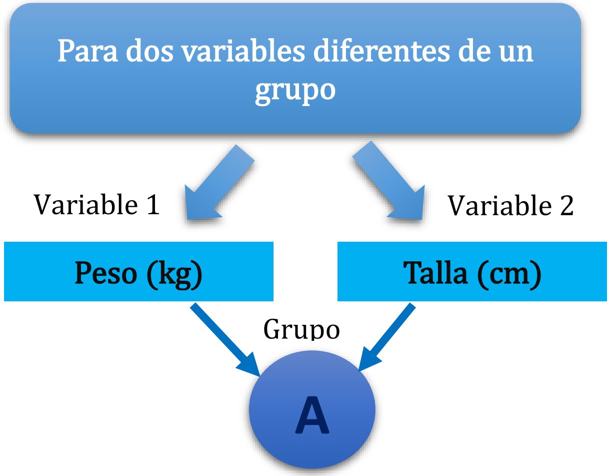
(Gálvez, Chirinos & Auqui, 2021)

Los siguientes esquemas podrían considerarse al momento de desarrollar una investigación.

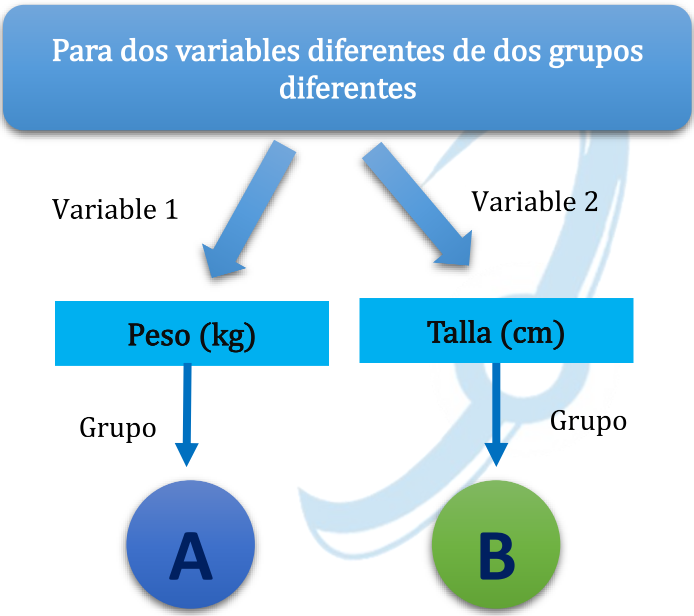
Primera situación:



Segunda situación:

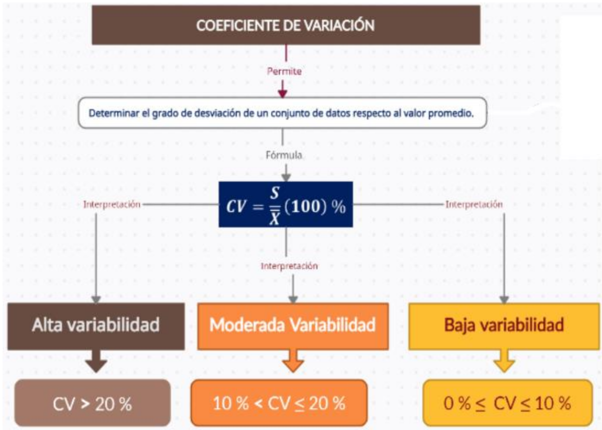


Tercera situación:



(Gálvez, Chirinos & Auqui, 2021)

Además, puede suceder todas las situaciones de comparación de los coeficientes de variación.



(Gálvez, Chirinos & Auqui, 2021)

(Gálvez, Chirinos & Auqui, 2021)

Ejemplo 15. En una encuesta a 30 estudiantes sobre la edad que tenía su madre cuando ellos nacieron, se registraron los siguientes datos:

28	30	22	27	30	25	31	35	24	30	24	28	34	36	28
28	31	35	28	30	21	31	23	29	30	24	28	30	20	32

Construya una distribución de frecuencias usando cuatro intervalos de clase y calcule la varianza, desviación estándar y el coeficiente de variación.

Solución:

Se halla el rango de los datos: R = 36 – 20 = 16. Como nos piden usar $k = 4$ intervalos de clase, luego, la amplitud de los intervalos es:

$$A = \frac{R}{k} = \frac{16}{4} = 4$$

Luego, construimos la tabla de frecuencias y hallamos las medidas pedidas.

Clases	x_i	f_i	$f_i . x_i$	$f_i . x_i^2$
[20, 24⟩	22	4	88	1936
[24, 28⟩	26	5	130	3380
[28, 32⟩	30	16	480	14400
[32, 36]	34	5	170	5780
Total		n = 30	868	25496

$$\bar{X} = \frac{868}{30} = 28{,}93 \text{ años}$$

Varianza:

$$\sigma^2 = \frac{\sum_{i=1}^{4} f_i\, x_i^2}{n} - \bar{X}^2 = \frac{25496}{30} - (28{,}93)^2$$

$$\sigma^2 = 849{,}8667 - 836{,}9449 = 12{,}9218 \text{ años}^2$$

Desviación estándar:

$$\sigma = \sqrt{\sigma^2} = \sqrt{12{,}9218} = 3{,}5947 \text{ años.}$$

Coeficiente de variación:

$$CV = \frac{\sigma}{\bar{X}}.100\ \% = \left(\frac{3{,}5947}{28{,}93}\right).100\ \% = 12{,}43\ \%$$

Ejemplo 16. Si en una distribución de frecuencias, la media es 16 y la varianza es 6,25; entonces el coeficiente de variación es

$$CV = \frac{\sigma}{\bar{X}}.100\ \% = \left(\frac{2{,}5}{16}\right).100\ \% = 15{,}63\ \%$$

donde, $\sigma = \sqrt{6{,}25} = 2{,}5$.

Nota. A menudo estaremos interesados en comparar la variabilidad de dos o más conjuntos de datos. Entonces,

a) Si dos muestras o poblaciones tienen varianzas (o desviaciones estándar) iguales, diremos que es “mejor” la que tiene mayor promedio. Si, por el contrario, las varianzas (o desviaciones estándar) son diferentes, diremos que es “mejor” la que tiene menor coeficiente de variación.

b) Si dos muestras o poblaciones tienen medias aritméticas iguales, diremos que es más "homogénea" la que tiene menor varianza (o menor desviación típica). Si las medias no son iguales, diremos que es más "homogénea" la que tiene menor coeficiente de variación.

Ejemplo 17. En una institución educativa se usa un método especial para mejorar el aprendizaje de los contenidos correspondiente al área de lógico – matemática. Al final del periodo académico se selecciona una muestra aleatoria de tamaño 20 de la institución educativa que le llamaremos grupo experimental. Paralelamente, se extrae una muestra aleatoria de tamaño 30 de una institución educativa que no está considerada como grupo experimental (grupo control). Se administra el mismo examen a los estudiantes elegidos como muestras, con los siguientes resultados:

Grupo experimental	12	09	15	13	10	13	15	11	14	13
	15	14	13	14	13	14	12	18	12	13
Grupo control	14	09	10	16	07	16	08	11	12	11
	12	14	14	09	12	11	12	08	15	11
	09	11	12	06	11	13	09	12	05	10

¿La metodología especial aplicada en el grupo experimental, mejora el aprendizaje?

Solución:

Para responder a la interrogante, se debe calcular la varianza, desviación estándar y el coeficiente de variación. Luego, se comparan los resultados.

a) Grupo experimental.

x_i	f_i	$f_i.x_i$	$f_i.x_i^2$
9	1	9	81
10	1	10	100
11	1	11	121
12	3	36	432
13	6	78	1014
14	4	56	784
15	3	45	675
18	1	18	324
Total	$n = 20$	263	3531

- $\bar{X} = \frac{263}{20} = 13{,}15$
- $S^2 = \frac{3531}{20} - (13{,}15)^2 \Rightarrow S^2 = 3{,}6275$
- $S = \sqrt{3{,}6275} = 1{,}9046 \Rightarrow S = 1{,}9046$
- $CV = \left(\frac{1{,}9046}{13{,}15}\right).100\ \% = 14{,}48\ \%$

b) Grupo Control.

x_i	f_i	$f_i.x_i$	$f_i.x_i^2$
5	1	5	25
6	1	6	36
7	1	7	49
8	2	16	128
9	4	36	324
10	2	20	200
11	6	66	726
12	6	72	864
13	1	13	169
14	3	42	588
15	1	15	225
16	2	32	512
Total	$n = 30$	330	3846

- $\bar{X} = \frac{330}{30} = 11$
- $S^2 = \frac{3846}{30} - (11)^2 \Rightarrow S^2 = 7{,}2$
- $S = \sqrt{7{,}2} = 2{,}6833 \Rightarrow S = 2{,}6833$
- $CV = \left(\frac{2{,}6833}{11}\right).(100)\% = 24{,}39\ \%$

Como las varianzas de los dos grupos son diferentes y como el coeficiente de variación del grupo experimental es menor que el coeficiente de variación del grupo control, entonces concluimos que el método aplicado mejora el aprendizaje de los contenidos del área lógica – matemática. Además, como las medias de los grupos son diferentes; las calificaciones del grupo piloto son más “homogéneas” que la del grupo control.

CAPÍTULO III

Coeficiente de Asimetría

Como las medidas de tendencia central, existen varias medidas de asimetría, una de ellas es el coeficiente o índice de asimetría de Pearson.

Definición. – El coeficiente de asimetría, denotado por A_s, mide el grado de deformación horizontal y se define por el número

$$A_s = \frac{3\,(\bar{X} - M_e)}{S}$$

donde, $\bar{X}$ es la media aritmética, M_e *e* es la mediana y S es la desviación estándar.

Al coeficiente de asimetría (A_s), se le conoce también como coeficiente de asimetría de Pearson; y cumple:

a) Si $A_s > 0$, la distribución es asimétrica positiva o sesgada a la derecha. En este caso, el orden de las medidas de tendencia central es: $\bar{X} > M_e > M_o$.

b) Si $A_s < 0$, la distribución es asimétrica negativa o sesgada a la izquierda. Aquí el orden de las medidas de tendencia central es: $\bar{X} < M_e < M_o$.

c) Si $A_s = 0$, la distribución es simétrica. En este caso coinciden las tres medidas de tendencia central: $\bar{X} = M_e = M_o$.

Ejemplo 18. Con los datos que se presenta en la siguiente tabla, calcular el grado de asimetría de la distribución y la magnitud del sesgo.

x_i	f_i	$f_i.x_i$	$f_i.x_i^2$	F_i
9	1	9	81	1
10	1	10	100	2
11	1	11	121	3
12	3	36	432	6
13	6	78	1014	12
14	4	56	784	16
15	3	45	675	19
18	1	18	324	20
Total	$n = 20$	263	3531	

Mediana

Solución:

Utilizando los procedimientos para calcular las medidas de tendencia central para datos agrupados, el valor de la media ($\bar{X}$)= 13,15 y la desviación estándar (S) = 1,9046). Falta calcular la mediana.

$\frac{n}{2} = \frac{20}{2} = 10.$ Entonces $M_e = 13$

Luego, $A_s = \frac{3\,(\bar{X} - M_e)}{S} = \frac{3\,(13{,}15-13)}{1{,}9046} = 0{,}2363$

Por tanto,

a) Como $A_s > 0$, la distribución es sesgada a la derecha.

b) La magnitud del sesgo es 0,2363.

Actividades

1. Un grupo de personas contestó a la pregunta, ¿cuántas películas ha visto en la última semana? Las respuestas fueron las siguientes:

4	7	3	2	0	2	1	2	3	0	1	4	3	3	3

Calcular la varianza, la desviación estándar y el coeficiente de variación.

2. Calcula la desviación estándar, para las tallas en cm, de los jugadores del equipo de baloncesto de una institución educativa

178	180	185	190	175	188	179	184

3. Observa las siguientes tablas de distribución de frecuencias. En cada caso completa y calcula la varianza, desviación estándar, el coeficiente de variación y el coeficiente de asimetría. ¿En cuál de las dos tablas los datos son más homogéneos?

Clase	f_i
[72; 78⟩	2
[78; 84⟩	5
[84; 90⟩	9
[90; 96⟩	11
[96; 102⟩	13
[102; 108⟩	6
[108; 114]	2

Clase	f_i
[18; 21 ⟩	3
[21; 24 ⟩	2
[; ⟩	8
[; ⟩	12
[; ⟩	11
[; ⟩	5
[;]	3

4. Una prueba de conocimientos, A, se calificó sobre 20 puntos dando una media de 12 y una desviación estándar de 2 puntos. Mientras que una prueba de aptitud, B, se calificó sobre 100 puntos, dando una media de 70 y una desviación estándar de 5.

   a) ¿En cuál de las dos pruebas los puntajes son más homogéneos?

      Respuesta: En la segunda prueba

   b) ¿Si Juan tiene 14 en A y Luís 73 en B, ¿quién tiene mejor rendimiento?

      Respuesta: Luís

5. En una prueba de aptitud, la distribución de frecuencias de las calificaciones resulta ser de forma unimodal con:

$\bar{X} = 90$, $M_e = 95$, y el coeficiente de asimetría $A_s = -0{,}75$. ¿Se puede afirmar que el coeficiente de variación de tales calificaciones es menor que 25 %?

Respuesta: Sí

Situación 1:

Se aplicó una encuesta a un grupo de hogares de cierto distrito en la que se analiza principalmente la información respecto al número de hijos mayores de 13 años que hay en el hogar y la cantidad de veces que éstos hayan tenido la necesidad de salir de casa en tiempos de pandemia para hacer algunas compras. Los datos obtenidos se muestran a continuación. ¿Cuál de los grupos de datos es menos disperso?

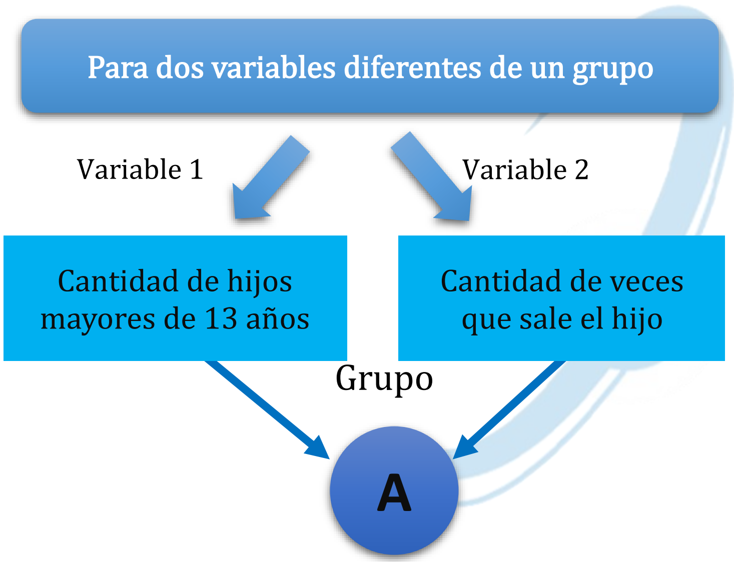
Cantidad de hijos mayores de 13 años				
1	3	2	3	4
4	2	0	4	4
1	3	4	1	3
2	5	2	4	4
4	1	3	3	1
2	1	4	2	4

Tipo de variable según su naturaleza: Cuantitativa discreta.
Variable: Cantidad de hijos mayores de 13 años en el hogar.

Cantidad de veces que ha salido el hijo del hogar				
10	12	11	11	13
15	16	0	12	13
11	12	11	13	13
13	12	16	15	11
12	16	10	12	16
17	11	13	12	10

Tipo de variable según su naturaleza: Cuantitativa discreta.
Variable: Cantidad de veces que sale un hijo de casa.

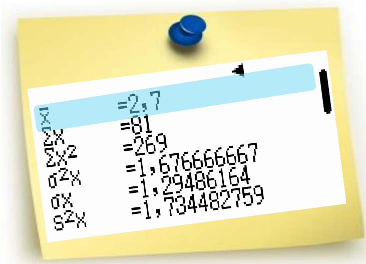
Se presenta la siguiente forma:

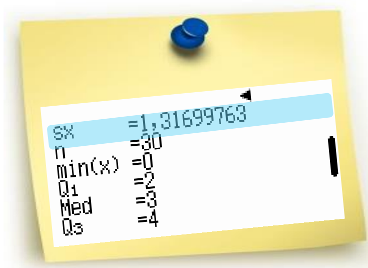


(Gálvez, Chirinos & Auqui, 2021)

Solución para la variable: Cantidad de hijos mayores de 13 años:

Se procede a realizar los cálculos obteniendo los resultados siguientes:





La media:

$\bar{X} = 2,7$

S=1,32

El coeficiente de variación:

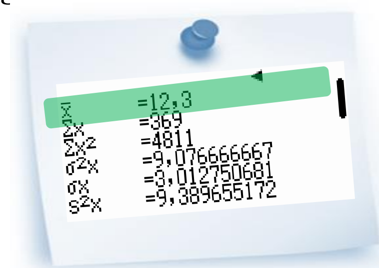
$$CV = \frac{S}{\bar{X}} = \frac{1,32}{2,7}\ (100)\ \%$$

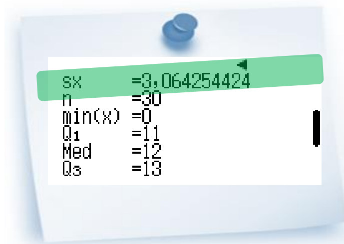
$$CV = 48,89\ \%$$

Los datos presentan “alta” variabilidad.

Solución para la variable: Cantidad de veces que ha salido el hijo del hogar.

Se procede a realizar los cálculos obteniendo los resultados siguientes:





La media:
$\bar{X} = 12,3$

S=3,06

El coeficiente de variación:

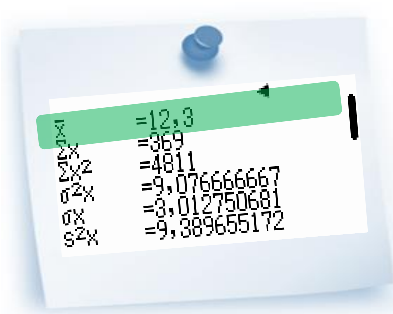
$$CV = \frac{S}{\bar{X}} = \frac{3,06}{12,3}\ (100)\ \%$$

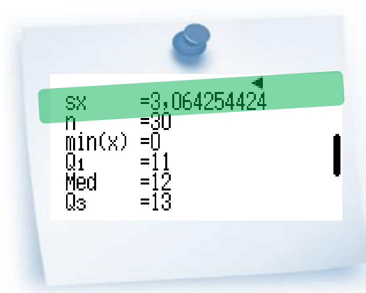
$$CV = 24,88\ \%$$

Los datos presentan "alta" variabilidad.

Solución para la variable: Cantidad de veces que ha salido el hijo del hogar.

Se procede a realizar los cálculos obteniendo los resultados siguientes:





La media:
$\bar{X} = 12,3$

S=3,06

El coeficiente de variación:

$$CV = \frac{S}{\bar{X}} = \frac{3,06}{12,3}\ (100)\ \%$$

$$CV = 24,88\ \%$$

Los datos presentan "alta" variabilidad.

Respuesta: Debido a que $\boldsymbol{CV_{N^\circ\ hijos} > CV_{N^\circ\ salidas}}$, entonces los datos que corresponden al número de salidas del hijo mayor de 13 años, es menos disperso que la otra variable.

Situación 2:

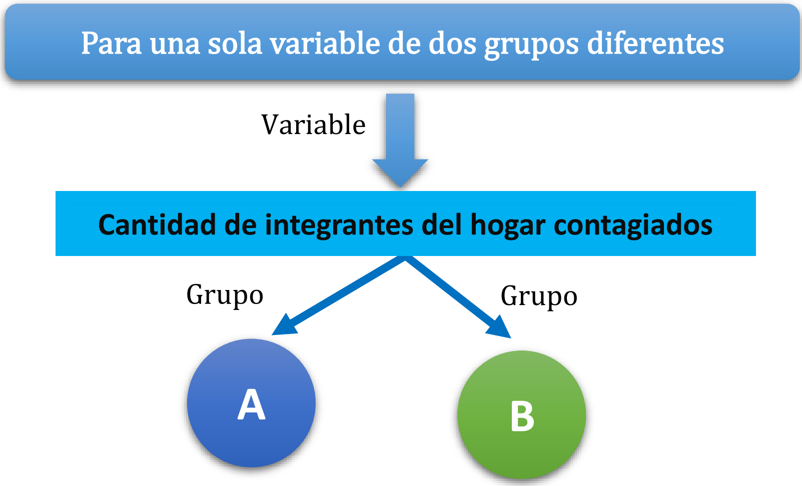
Un grupo de investigadores han aplicado una encuesta a 200 familias de Chosica y Huaycán para saber sobre el número de integrantes que se hayan contagiado con la COVID-19. Los resultados se muestran a continuación:

Distribución de familias según la cantidad de integrantes contagiados con la COVID-19

N° de contagiados	Chosica	Huaycán
0	52	10
1	45	85
2	58	62
3	29	35
4	6	2
5	1	1
6	9	5
Total	200	200

¿Cuál de los dos distritos es más disperso respecto al número de integrantes de la familia contagiados con la COVID-19?

Se presenta la siguiente forma:



Solución:

Tabla de distribución de frecuencia de familias de Huaycán según la cantidad de integrantes contagiados con la COVID-19

x_i	f_i	$x_i . f_i$
0	10	0
1	85	85
2	62	124
3	35	105
4	2	8
5	1	5
6	5	30
Total	200	357

La media:

$$\bar{X} = \frac{357}{200} = 1{,}785$$

S=1,12

El coeficiente de variación:

$$CV = \frac{S}{\bar{X}} = \frac{1{,}12}{1{,}785}\ (100)\ \%$$

$$CV = 62{,}75\ \%$$

Los datos presentan "alta" variabilidad.

Distribución de familias de Chosica según la cantidad de integrantes contagiados con la COVID-19

x_i	f_i	$x_i . f_i$
0	10	0
1	85	85
2	62	124
3	35	105
4	2	8
5	1	5
6	5	30
Total	200	357

La media:

$$\bar{X} = \frac{331}{200} = 1,655$$

S=1,479

El coeficiente de variación:

$$CV = \frac{S}{\bar{X}}$$

Los datos presentan "alta" variabilidad.

Respuesta: Debido a que $\boldsymbol{CV_{Chosica}} > \boldsymbol{CV_{Huaycán}}$, entonces el contagio de familias de Chosica es más disperso que de Huaycán.

Situación 3:

La siguiente tabla de distribución de frecuencias muestra datos que corresponden a los pesos en kilogramos de 50 pacientes recuperados de la COVID-19 que asistieron a sus citas de nutrición el mes pasado.

Distribución de pacientes de nutrición según su peso en kilogramos

Peso (kg)	f_i
[35; 45⟩	2
[45; 55⟩	2
[55; 65⟩	7
[65; 75⟩	13
[75; 85⟩	11
[85; 95⟩	11
[95; 105]	4
Total	$n = 50$

Calcule e interprete el coeficiente de variación.

Se presenta la siguiente forma:

Para una variable de un grupo

↓ Variable 1

Peso de pacientes recuperados de la COVID - 19

↓ Grupo

A

Distribución de frecuencias de pacientes de nutrición según su peso en kilogramos.

Peso (kg)	$\boldsymbol{f_i}$	$\boldsymbol{\dot{x}_i}$	$\boldsymbol{\dot{x}_i . f_i}$
[35; 45⟩	2	40	80
[45; 55⟩	2	50	100
[55; 65⟩	7	60	420
[65; 75⟩	13	70	910
[75; 85⟩	11	80	880
[85; 95⟩	11	90	990
[95; 105]	4	100	400
Total	$n = 50$		3780

Media:

$$\bar{X} = \frac{3780}{50}$$

$$\bar{X} = 75{,}6$$

La desviación estándar:

$$S = 14{,}87$$

Coeficiente de variación:

$$CV = \frac{14{,}87}{75{,}6}(100)\ \%$$

$$CV = 19{,}67\ \%$$

NOTA: Considerar que cuando se trabaja con intervalos en la calculadora, el cálculo de la desviación estándar se hará ingresando la marca de clase y la frecuencia absoluta simple.

Interpretación: El peso de los 50 pacientes de nutrición recuperados de la COVID-19 tiene moderada variabilidad.

Referencias

Gálvez, L., Chirinos, D. M.& Auqui, E. K. (2021). *Situaciones didácticas para la enseñanza de la media, desviación estándar y coeficiente de variación en estudiantes de educación superior utilizando el emulador de la calculadora científica Casio fx-570LA X Classwiz*. RELME 35. República Dominicana. Recuperado 20-09-2022 de Free Math Teaching Materials | CASIO (casio-intl.com)

Godino, J. D., Batanero, C., & Roa, R. (2005). *An onto-semiotic analysis of combinatorial problems and the solving processes by university students*. *Educational Studies in Mathematics*, *60*(1), 3–36.

Hanke, H. E. y Reitsh, A. (1996). Pronostico en los negocios. 5ed. México. Pearson Educación.

Mayorga-Ponce, R. B., Reyes-Torres, S. B., Baltazar-Téllez, R. M., & Martínez-Alamilla, A. (2021). Medidas de Dispersión. *Educación Y Salud Boletín Científico Instituto De Ciencias De La Salud Universidad*

Autónoma Del Estado De Hidalgo, 9 (18), 77-79. https://doi.org/10.29057/icsa.v9i18.7115

Molina, G. & Rodrigo, M. F (2010). *Estadísticos de dispersión*. Open curso ware recuperado el 19-09-2022 de http://ocw.uv.es/ciencias-de-la-salud/pruebas-1/1-3/t_04.pdf.

Pagano, R.R. (1998). *Estadística para la ciencia del comportamiento*. 7ma. Editorial: Thomsom. España.

Vilches, M, A. (2014). *Cuartiles, deciles y percentiles*. Universidad Autónoma de Estado de Hidalgo. Sistema de Universidad Virtual. México. Obtenido de: https://www.academia.edu/36233311/_Cuartiles_deciles_y_percentiles_

Vílchez, J. (2013). *Elementos de estadística y probabilística*. Editorial Gráfica Carvil. S.A.C. Ate, Lima.