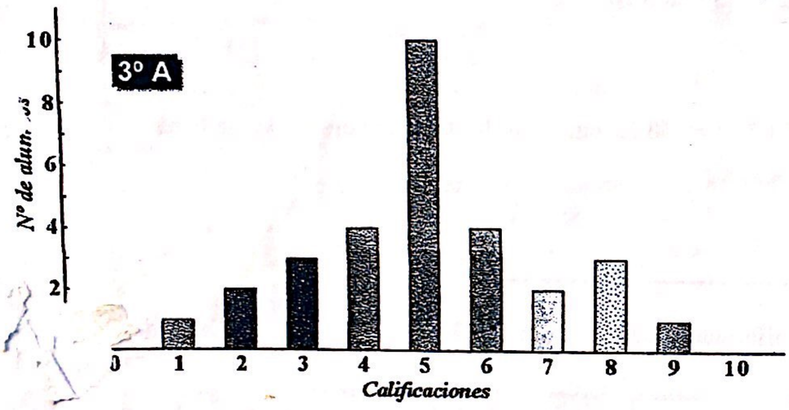
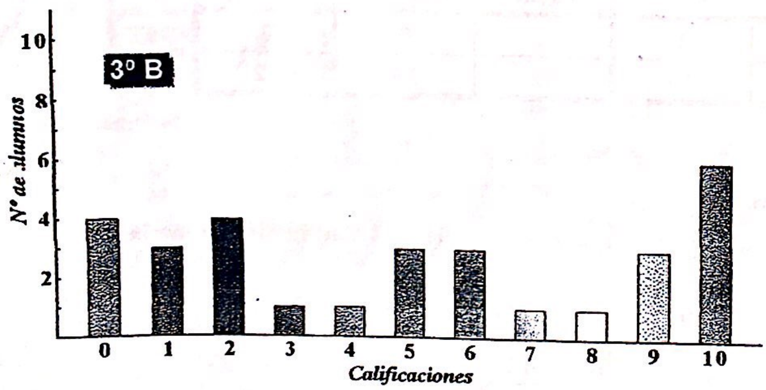
3. Parámetros de dispersión

Este año hay dos cursos muy desiguales en cuanto al rendimiento en Matemáticas. Observa sus calificaciones.



¡Poca dispersión!

En 3º A hay pocas notas bajas, pocas altas y casi todas se sitúan en torno al 5.



¡Mucha dispersión!

En 3º B hay bastantes alumnos con muy bajo rendimiento, bastantes con muy buen rendimiento y pocas calificaciones en torno al 5.

Aunque estas distribuciones de notas tienen aspecto diferente, sus medias son parecidas, $\bar{x}_A = 5'03$ y $\bar{x}_B = 5'1$; es decir, lo que diferencia a ambos cursos es su comportamiento respecto a la media.

Es necesario, pues, conocer en qué medida los datos numéricos están agrupados o no alrededor de los valores centrales. A esto es a lo que se llama ***dispersión,*** y los parámetros que nos informan de cómo se separan los datos se llaman *parámetros* o *medidas de dispersión.*

Los ***parámetros de dispersión*** son valores numéricos que nos informan de las desviaciones que sufren los datos de una distribución estadística respecto de los parámetros centrales, en particular respecto a la media aritmética.

3.1. Rango o recorrido

Una manera muy sencilla de determinar el grado de dispersión de los datos es observar la separación entre el dato más grande y el más pequeño de la distribución estadística.

Rango o ***recorrido*** de una variable estadística es la diferencia entre el mayor y el menor valor de la variable estadística. Se representa por ***R.***

Ejemplo.- Halla el rango de la siguiente distribución estadística.

Puntuación	[38, 44)	[44, 50)	[50, 56)	[56, 62)	[62, 68)	[68, 74)	[74, 80]
Nº de alumnos	7	8	15	25	18	9	6

$R = 80 - 38 = 42$ puntos

El rango es un parámetro fácil de calcular, pero que ofrece una información muy limitada. Así, nos da una idea de la amplitud del conjunto de datos, pero está muy influido por los valores extremos.

3.2. Desviación media

La distancia entre cualquier dato y la media aritmética, $|x_i - \bar{x}|$, recibe el nombre de *desviación* de dicho dato. Una manera de observar la dispersión de la distribución estadística es calcular la media aritmética de todas las desviaciones.

Desviación media de una variable estadística es la media aritmética de las desviaciones de todos los datos respecto a su media aritmética. Se representa por d_m.

- Si la variable toma los N valores $x_1, x_2, \ldots, x_N$ (datos sin frecuencia) la desviación media se puede calcular mediante la siguiente expresión:

$$d_m = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + \ldots + |x_N - \bar{x}|}{N} = \frac{\sum_{i=1}^{N} |x_i - \bar{x}|}{N}$$

- Si la variable toma los valores o marcas de clase $x_1, x_2, \ldots, x_k$, con frecuencias absolutas $n_1, n_2, \ldots, n_k$, la desviación media se calcula con la expresión siguiente:

$$d_m = \frac{n_1|x_1 - \bar{x}| + n_2|x_2 - \bar{x}| + \ldots + n_k|x_k - \bar{x}|}{N} = \frac{\sum_{i=1}^{k} n_i |x_i - \bar{x}|}{N}$$

Ejemplo.- Los siguientes datos corresponden al número de faltas de ortografía cometidas por dos alumnos en siete dictados.

Alumno A: 1, 2, 5, 5, 5, 8, 9 Alumno B: 4, 4, 5, 5, 5, 6, 6

Hallamos la desviación media de ambas series de datos:

$$\bar{x}_A = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{1+2+5+5+5+8+9}{7} = \frac{35}{7} = 5 \text{ faltas}$$

$$d_{mA} = \frac{\sum_{i=1}^{N} |x_i - \bar{x}|}{N} = \frac{|1-5|+|2-5|+|5-5|+|5-5|+|5-5|+|8-5|+|9-5|}{7} = \frac{14}{7} = 2 \text{ faltas}$$

Para la otra serie, agrupamos los datos en una tabla:

x_i	n_i	$n_i x_i$	$\|x_i - \bar{x}\|$	$n_i \|x_i - \bar{x}\|$
4	2	8	1	2
5	3	15	0	0
6	2	12	1	2
Total	7	35		4

$$\bar{x}_B = \frac{\sum_{i=1}^{k} n_i x_i}{N} = \frac{35}{7} = 5 \text{ faltas} \Rightarrow d_{mB} = \frac{\sum_{i=1}^{k} n_i |x_i - \bar{x}|}{N} = \frac{4}{7} = 0'5714 \text{ faltas}$$

Observamos que, aunque ambos tienen igual media aritmética, el número de faltas de ortografía está menos disperso en el segundo alumno (es decir, el alumno B es mucho más regular que el alumno A).

3.3. Varianza y desviación típica

Otro parámetro estadístico importante es el que mide la dispersión a partir de los cuadrados de las desviaciones.

Varianza de una variable estadística es la media aritmética de los cuadrados de las desviaciones de todos los datos respecto a su media aritmética. Se representa por σ^2.

- Si la variable toma los N valores $x_1, x_2, \ldots, x_N$ (datos sin frecuencia) la varianza se puede calcular mediante alguna de las siguientes expresiones:

$$\sigma^2 = \frac{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + \ldots + (x_N-\bar{x})^2}{N} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})^2}{N}$$

o bien mediante $$\sigma^2 = \frac{x_1^2 + x_2^2 + \ldots + x_N^2}{N} - \bar{x}^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - \bar{x}^2$$

- Si la variable toma los valores o marcas de clase $x_1, x_2, \ldots, x_k$, con frecuencias absolutas $n_1, n_2, \ldots, n_k$, la varianza se calcula mediante las expresiones siguientes:

$$\sigma^2 = \frac{n_1(x_1-\bar{x})^2 + n_2(x_2-\bar{x})^2 + \ldots + n_k(x_k-\bar{x})^2}{N} = \frac{\sum_{i=1}^{k} n_i(x_i-\bar{x})^2}{N}$$

Escaneado con CamScanner

o bien mediante $\sigma^2 = \frac{n_1 x_1^2 + n_2 x_2^2 + \ldots + n_k x_k^2}{N} - \overline{x}^2 = \frac{\sum_{i=1}^{k} n_i x_i^2}{N} - \overline{x}^2$

Desviación típica de una variable estadística es la raíz cuadrada positiva de la varianza. Se denota por σ.

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

3.4. Coeficiente de variación

Una dispersión de un metro en la medida de una longitud de 100 km es muchísimo más pequeña que una dispersión de un metro en una medida de 10 km.

El cociente entre la desviación típica y la media de una variable estadística se denomina ***coeficiente de variación*** y es muy útil para comparar las dispersiones de dos variables estadísticas de diferente media o de diferente naturaleza. Se suele expresar en % y se representa por C_{var}.

$$C_{var} = \frac{\sigma}{\overline{x}}$$

La media $\overline{x}$, así como la desviación típica σ se expresan en la misma unidad que la variable X. El coeficiente de variación es una cantidad sin dimensión, independientemente de las unidades elegidas.

Ejemplo.- ¿Qué serie de números te parece más dispersa de las siguientes?

Serie A: 1, 3, 5, 7, 9 Serie B: 1, 4, 8, 8

Hallamos la media y la desviación típica de ambas series para calcular sus coeficientes de variación:

Serie A: $\overline{x}_A = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{1+3+5+7+9}{5} = \frac{25}{5} = 5$

$$\sigma_A^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - \overline{x}^2 = \frac{1^2+3^2+5^2+7^2+9^2}{5} - 5^2 = \frac{165}{5} - 25 = 8\,;$$

luego $\sigma_A = \sqrt{8} \cong 2'8284$

Serie B: $\overline{x}_B = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{1+4+8+8}{4} = \frac{21}{4} = 5'25$

$$\sigma_B^2 = \frac{\sum_{i=1}^{N} x_i^2}{N} - \overline{x}^2 = \frac{1^2+4^2+8^2+8^2}{4} - 5'25^2 = \frac{145}{4} - 27'5625 = 8'6875\,;$$

con lo que $\sigma_B = \sqrt{8'6875} \cong 2'9475$

Con los datos anteriores, los coeficientes de variación de las respectivas series son:

$$C_{\text{var}A} = \frac{\sigma_A}{\overline{x}_A} = \frac{2'8284}{5} \cdot 100 = 56'57\ \% \qquad C_{\text{var}B} = \frac{\sigma_B}{\overline{x}_B} = \frac{2'9475}{5'25} \cdot 100 = 56'14\ \%$$

Por tanto, la segunda serie es algo menos dispersa que la primera (aunque tenga mayor desviación típica).

Ejemplo.- Analicemos los parámetros de dispersión de las distribuciones estadísticas vistas anteriormente relativas a las calificaciones de los cursos 3º A y 3º B.

- **Distribución estadística de 3º B**

x_i	n_i	$n_i x_i$	$\lvert x_i-\bar{x}\rvert$	$n_i\lvert x_i-\bar{x}\rvert$	$(x_i-\bar{x})^2$	$n_i(x_i-\bar{x})^2$
0	4	0	5'1	20'4	26'01	104'04
1	3	3	4'1	12'3	16'81	50'43
2	4	8	3'1	12'4	9'61	38'44
3	1	3	2'1	2'1	4'41	4'41
4	1	4	1'1	1'1	1'21	1'21
5	3	15	0'1	0'3	0'01	0'03
6	3	18	0'9	2'7	0'81	2'43
7	1	7	1'9	1'9	3'61	3'61
8	1	8	2'9	2'9	8'41	8'41
9	3	27	3'9	11'7	15'21	45'63
10	6	60	4'9	29'4	24'01	144'06
Total	**30**	**153**		**97'2**		**402'7**

$$\bar{x}_B = \frac{\sum_{i=1}^{k} n_i x_i}{N} = \frac{153}{30} = 5'1 \text{ puntos} \qquad \sigma_B^2 = \frac{\sum_{i=1}^{k} n_i (x_i-\bar{x})^2}{N} = \frac{402'7}{30} = 13'42$$

$$R_B = 10 - 0 = 10 \text{ puntos} \qquad \sigma_B = \sqrt{\sigma^2} = \sqrt{13'42} = 3'66 \text{ puntos}$$

$$d_{mB} = \frac{\sum_{i=1}^{k} n_i \lvert x_i-\bar{x}\rvert}{N} = \frac{97'2}{30} = 3'24 \text{ puntos} \qquad C_{\text{var}B} = \frac{\sigma_B}{\bar{x}_B} = \frac{3'66}{5'1}\cdot 100 = 71'76\ \%$$

- **Distribución estadística de 3º A**

x_i	n_i	$n_i x_i$	$\lvert x_i-\bar{x}\rvert$	$n_i\lvert x_i-\bar{x}\rvert$	$n_i x_i^2$
1	1	1	4'03	4'03	1
2	2	4	3'03	6'06	8
3	3	9	2'03	6'09	27
4	4	16	1'03	4'12	64
5	10	50	0'03	0'30	250
6	4	24	0'97	3'88	144
7	2	14	1'97	3'94	98
8	3	24	2'97	8'91	192
9	1	9	3'97	3'97	81
Total	**30**	**151**		**41'30**	**865**

$$\bar{x}_A = \frac{\sum_{i=1}^{k} n_i x_i}{N} = \frac{151}{30} = 5'03 \text{ puntos} \qquad \sigma_A^2 = \frac{\sum_{i=1}^{k} n_i x_i^2}{N} - \bar{x}^2 = \frac{865}{30} - 5'03^2 = 3'53$$

$$R_A = 9 - 1 = 8 \text{ puntos} \qquad \sigma_A = \sqrt{\sigma^2} = \sqrt{3'53} = 1'89 \text{ puntos}$$

$$d_{mA} = \frac{\sum_{i=1}^{k} n_i \lvert x_i-\bar{x}\rvert}{N} = \frac{41'30}{30} = 1'38 \text{ puntos} \qquad C_{\text{var}A} = \frac{\sigma_A}{\bar{x}_A} = \frac{1'89}{5'03}\cdot 100 = 37'57\ \%$$

Conclusiones:

Parámetros de dispersión	*3º A*	*3º B*
Rango	8	10
Desviación media	1'38	3'24
Desviación típica	1'89	3'66
Coeficiente de variación	37'57 %	71'76 %

Como podemos apreciar, todos los parámetros de dispersión del grupo 3º A son menores que los del grupo 3º B, incluido el coeficiente de variación. Por tanto, podemos afirmar rotundamente que la distribución estadística del grupo 3º A es menos dispersa que la de 3º B.

EJERCICIOS

9. Las calificaciones de Juan en seis pruebas fueron: 87, 64, 92, 86, 69 y 71. Halla la media, la mediana y todos los parámetros de dispersión.

10. Fíjate que para hallar la varianza hay que elevar al cuadrado las desviaciones respecto a la media; por ello, la varianza no se expresa en las mismas unidades que los datos. De manera que si los datos se expresan en metros, ¿en qué unidades se expresará la varianza? ¿Y la desviación típica y el coeficiente de variación?

11. Los siguientes datos son calificaciones obtenidas en cierto examen de Lengua.

    2, 5, 3, 4, 7, 9, 5, 2, 7, 4, 8, 3, 5, 8, 7, 9, 3, 2, 4, 1, 10, 9, 4, 8, 6, 9, 3, 3, 7, 1, 2, 8, 6, 7, 3, 6, 4, 7, 4, 8,
    2, 3, 7, 5, 4, 6, 7, 5, 6, 7, 8, 4, 3, 7, 5, 6, 9, 5, 7, 2

    a) Elabora una tabla en la que aparezcan las diferentes frecuencias simples.
    b) Calcula los parámetros de centralización de las calificaciones.
    c) Calcula todos los parámetros de dispersión.

12. En la fabricación de cierto tipo de bombillas se han detectado algunas defectuosas. Se han estudiado 200 lotes de 500 piezas cada uno, obteniéndose los datos de la tabla adjunta.

Defectuosas	1	2	3	4	5	6	7	8
Nº de lotes	5	15	38	42	49	32	17	2

Calcula los parámetros de centralización y de dispersión.

13. En un hospital se quiere estimar el peso de los niños recién nacidos. Para ello se seleccionan, de forma aleatoria, 100 de éstos, obteniéndose los siguientes resultados.

***Peso* (kg)**	[1, 1'5)	[1'5, 2)	[2, 2'5)	[2'5, 3)	[3, 3'5)	[3'5, 4)	[4, 4'5)	[4'5, 5]
Nº de niños	1	2	5	20	40	26	5	1

    a) Calcula los pesos medio, mediano y moda de la distribución anterior.
    b) Determina el rango, la desviación media y la desviación típica de la variable.

14. Si has realizado los ejercicios 12 y 13 anteriores podrás comprobar que las desviaciones típicas son, respectivamente, 1'5254 y 0'5679. ¿Cuál de las dos distribuciones es menos dispersa?

15. Si a los números 10, 12, 14, 16, 18 y 20, los multiplicamos por 4 se obtiene 40, 48, 56, 64, 72 y 80. ¿Qué puedes decir de las medias, las varianzas y las desviaciones típicas de ambas series estadísticas?

16. Si a los números 10, 12, 14, 16, 18 y 20, les sumamos 9 se obtiene 19, 21, 23, 25, 27 y 29. Compara las medias, las varianzas y las desviaciones típicas de ambas series estadísticas.