Unidad 5

Estadística

Cuando se escuchan noticias en la radio y en la televisión, se leen los periódicos o alguna revista, suele encontrarse gran cantidad de datos numéricos. Estos datos se pueden referir al desempleo, a los deportes, a la producción industrial, a la esperanza de vida o a otros temas de interés para los ciudadanos y para los que gobiernan y toman decisiones basadas en la información que poseen. Esos números se denominan *cifras estadísticas*. Los procedimientos para recoger, clasificar, resumir, analizar datos y elaborar conclusiones a partir de esa información son estudiados por la Estadística. Algunos problemas que afectan a gran número de personas (como el control de las enfermedades o el uso racional del agua potable) se pueden estudiar con la ayuda de estos procesos matemáticos cuyos resultados se muestran mediante distintas expresiones numéricas o gráficas.

La información que permite estudiar cómo se comporta alguna característica de cierta población se reúne en tablas o series estadísticas. En particular, las formas de representación que analizarás en esta unidad son útiles cuando se maneja gran cantidad de información a partir de la cual se pueden realizar deducciones o inferencias que resultan confiables.

A

1. Para saber lo que sabés

Antes de comenzar con el desarrollo de las actividades de esta unidad es conveniente que revises el significado de algunos términos que se usan en Estadística y que aprendiste en años anteriores.

Términos estadísticos

- **Población:** es el conjunto de individuos (personas o cosas) sobre los que se realiza una estadística.
- **Muestra:** es un conjunto de individuos que han sido seleccionados de una población y que la representan fielmente; el modo de elegir la muestra determina si será representativa o no del conjunto al que pertenece.
- **Variable estadística:** es una característica de la población que se quiere estudiar y que puede variar por algún motivo, es decir, que puede tener distintos valores.
- **Frecuencia absoluta:** es el número de veces que se presenta cada valor de la variable estadística.
- **Frecuencia relativa:** es la frecuencia expresada como un porcentaje con respecto al total de observaciones.

• • • Valores centrales o medidas de centralidad

• **Media o promedio:** es igual a la suma de todos los valores observados dividida por el número de observaciones realizadas.
• **Moda:** es el valor de la variable que tiene mayor frecuencia.
• **Mediana:** es el valor central que separa a las observaciones, previamente ordenadas en dos grupos de igual cantidad de datos cada uno.

*Si tenés alguna duda respecto del significado y uso de alguno de estos términos, consultá con tu docente acerca de la conveniencia de volver a revisar la unidad **5** del CUADERNO DE ESTUDIO **1** o de consultar alguno de los libros de la biblioteca de tu escuela.*

A

2. Histogramas

En esta actividad vas a confeccionar un gráfico estadístico formado por rectángulos. Este tipo de gráficos se suele utilizar cuando se quiere presentar datos que han sido organizados agrupándolos por intervalos.

En una escuela urbana, el profesor de Educación Física registró, en centímetros, la talla de sus 40 alumnos e hizo las siguientes anotaciones:

165	**160**	**163**	**175**	**174**	**160**	**165**	**154**	**168**	**165**
168	**168**	**158**	**162**	**160**	**161**	**162**	**166**	**163**	**159**
178	**169**	**178**	**169**	**171**	**170**	**168**	**150**	**167**	**168**
149	**165**	**168**	**156**	**175**	**168**	**173**	**172**	**163**	**164**

Como quería mostrar los resultados en un gráfico y las alturas registradas eran muy diversas, consideró conveniente agrupar primero los datos en intervalos de 5 centímetros de amplitud, aunque podría haber tomado intervalos más pequeños o más grandes.

a) Podés organizarte para trabajar con un compañero de la siguiente manera:

1. Copien en la carpeta una disposición de los intervalos como la siguiente:
entre 148,5 y 153,5: . . . . .Total: . . . .
entre 153,5 y 158,5: . . . . Total: . . . .
entre 158,2 y 163,5: . . . . .Total: . . . .
entre 163,3 y 168,5: . . . . .Total: . . . .
entre 168,5 y 173,5: . . . . .Total: . . . .
entre 173,5 y 178,5: . . . . .Total: . . . .

2. Mientras uno de ustedes va leyendo las alturas registradas, el otro hará una marca en el intervalo al que corresponde. Tengan en cuenta que si agrupan las marcas que van realizando en manojos de 5 después les resultará más fácil contar cuántas marcas corresponden a cada intervalo.
3. Luego de contarlas, escriban las frecuencias totales, es decir, cuántas alturas fueron registradas para cada intervalo.
4. A partir de esta información ya organizada, van a realizar el gráfico.

*Un **histograma** es un gráfico de barras que, para cada intervalo de la variable en estudio, muestra un rectángulo con base en una línea con los valores considerados y altura proporcional a la frecuencia.*

- Construyan, en papel cuadriculado, un histograma de seis rectángulos, uno por cada intervalo. En el **eje x** escriban los límites de los intervalos y en el **eje y** las frecuencias correspondientes. Peguen el histograma en sus carpetas.

b) En la unidad **5** del CUADERNO DE ESTUDIO **1** aprendiste a calcular los valores centrales de una serie de datos estadísticos. Respondé en tu carpeta cuál es la moda, la mediana y la media o promedio de los datos registrados por el profesor de Educación Física.

A 3. Estudio de la dispersión: varianza

Luego de revisar las medidas de centralidad, en esta actividad estudiarás cómo se puede medir la dispersión de los datos cuando en una estadística se encuentran muy alejados del promedio.

Tal como estudiaste en años anteriores, los *valores centrales* o *parámetros* (como la *moda*, la *mediana* y la *media o promedio*) resumen en un sólo número la serie de datos que forman parte de una distribución. Se considera que su utilidad es mayor cuanto mejor representan a todo el conjunto.

El *promedio* o *media aritmética* es el valor central más utilizado en los trabajos de Estadística. Esto se debe a que su significado puede interpretarse inmediatamente y, por otro lado, resulta muy sencillo calcularlo. Sin embargo, no siempre es el valor más representativo de los datos considerados. En algunos casos, el valor del promedio puede estar fuertemente influido por los valores extremos de la serie que se está estudiando. Esto ocurre cuando esos valores extremos están muy alejados del promedio.

Los siguientes ejemplos te permitirán comprender la relación entre los valores centrales y la utilidad que poseen.

En una familia donde la abuela tiene 78 años, la madre 40, el padre 45, la tía 43 y los hijos 5, 8 y 11; los extremos (que son 5 y 78 años) están muy alejados del promedio, que es de 33 años. En este caso el promedio de edades de los miembros de esa familia carece de representatividad.

Otro ejemplo. El alumno A obtiene las siguientes notas: 7; 8; 7; 9; 8; 9; 9; 9; 8, 9; mientras que las del alumno B son: 3, 8, 9, 1, 6, 7, 1, 8, 8. La media o promedio del alumno A es 8,30 y la del alumno B, 5,1. En el primer caso, la media proporciona buena información acerca del rendimiento escolar de A porque la mayoría de las notas están cercanas al promedio general; es decir, que los valores observados están muy concentrados alrededor del promedio, por lo tanto, es muy representativo de la situación de este alumno. En cambio, el promedio del alumno B no resulta representativo ya que el conjunto de notas que se busca representar es disperso.

Para que un valor central sea representativo de una población debe ser lo más próximo posible a las características de la población total que se está investigando.

Este es otro ejemplo. Si las estaturas, en metros, de los miembros de una familia son: 1,20; 1,32; 1,48; 1,65; 1,69; 1,85; el promedio es 1,53 pero no resulta muy representativo porque los valores encontrados para las estaturas distan mucho de ese promedio. Por lo tanto, no tiene sentido presentar el dato "promedio de alturas", ya que no brinda buena información sobre esta familia. En cambio, como viste en el caso de las alturas de los alumnos de una sección escolar, conocer el promedio de alturas de los chicos resulta de utilidad.

Para mejorar la representatividad de un parámetro, en toda distribución estadística, se debe considerar el promedio y la ***dispersión*** *de los datos con respecto a él. Esa dispersión o desviación de los datos es una* ***medida descriptiva*** *que indica la mayor o menor concentración de los datos con relación a un parámetro de centralización.*

a) Leé la siguiente tabla en la que se indican las temperaturas, en grados centígrados, tomadas en dos ciudades, C_1 y C_2, el día 15 de cada mes de un mismo año, a las 12 del mediodía.

	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto	Septiembre	Octubre	Noviembre	Diciembre
C_1	26	25	20	15	14	11	11	14	15	14	19	24
C_2	37	38	24	14	10	6	-2	6	12	13	20	28

1. Observá los datos de la tabla y calculá, para cada una de las ciudades, la diferencia entre el mayor y el menor valor. Esa diferencia se llama *rango* o *recorrido* de una distribución.
2. ¿En cuál de las dos ciudades el rango de los valores registrados es más amplio?
3. Calculá la temperatura media en las dos ciudades.
4. ¿Te parece que en ambas la media o promedio es igualmente representativo? Justificá tu respuesta.

*El **rango** o **recorrido** de una distribución de datos es la diferencia entre el mayor y el menor valor. Si el rango de una distribución es grande indica que existen valores muy alejados de la media.*

Habrás observado que si bien los dos promedios que obtuviste son parecidos (17,3° C para C_1 y 17,2° C para C_2) a lo largo del año las temperaturas son más moderadas en C_1 que en C_2, en la que son notablemente más extremas. Como el promedio no pone en evidencia esas diferencias respecto de la media, será necesario buscar un procedimiento más adecuado que permita conocer esa información.

b) Construí en tu carpeta una tabla como la siguiente. Colocá en cada casilla la diferencia entre el valor registrado y el promedio, como se muestra en los ejemplos:

	C_1	C_2
Enero	26 -17,3 = 8,7	37 - 17,2 = 19,8
Febrero		
Marzo		
Abril		
Mayo		
Junio		
Julio	11-17,3= -6,3	-2-17,2= -19,2
Agosto		
Septiembre		
Octubre		
Noviembre		
Diciembre		

La diferencia entre el valor registrado y el promedio te permitió conocer a qué distancia del promedio está cada uno de los datos.

*En estadística, se llama **dispersión** a la distancia entre los datos y el promedio.*

Como habrás podido observar, algunas diferencias tienen signo positivo y otras, negativo. Podría suceder que las diferencias de distinto signo se compensaran y entonces la suma de las desviaciones resultara nula. Para evitar eso, se considera el cuadrado de cada una de las desviaciones ya que los cuadrados son siempre positivos.

Por ejemplo: $(37 - 17{,}2)^2 = 19{,}8^2 = 392{,}04$.
$(-2 - 17{,}2)^2 = (-19{,}2)^2 = 368{,}64$.

En estadística se usa el símbolo $\bar{x}$ para representar a la media o promedio. En símbolos, el procedimiento anterior se puede entonces generalizar escribiendo:

• $(x_i - \bar{x})$ que simboliza la diferencia entre un valor y el promedio; el subíndice i representa al número que indica cuál es la posición en la tabla del valor x que se está considerando.

• $(x_i - \bar{x})^2$ es el cuadrado de la diferencia entre el valor considerado y el promedio.

• El símbolo Σ (que es la letra griega sigma mayúscula) se usa para representar la suma de *n* términos que se puede calcular mediante una fórmula.

Para estimar la dispersión de un conjunto de datos con relación al promedio, en estadística se define la varianza.

*Se llama **varianza** a la suma de los cuadrados de todas las diferencias $(x_i - \bar{x})$ desde x_1 hasta x_n dividida por el número de términos.*
*Se simboliza: **Varianza** $= \dfrac{\sum_{1}^{n}(x_i - \bar{x})^2}{n}$*

Observá que en la expresión simbólica anterior el signo Σ, tiene un subíndice y un exponente que indican los sucesivos valores que toma el subíndice **i** de **x** en cada uno de los términos de la sumatoria.

c) Para calcular la varianza en el caso de las temperaturas de las ciudades C_1 y C_2 con las que trabajaron en la consigna **b** sigan los siguientes pasos:

1. Calculen para cada ciudad el cuadrado de las diferencias que obtuvieron en la consigna **b** y completen una tabla como la siguiente:

	C_1		C_2	
	$(x_i - x)$	$(x_i - \bar{x})^2$	$(x_i - x)$	$(x_i - \bar{x})^2$
Enero 1	8,7	75,69	19,8	392,04
Febrero 2				
Marzo 3				
Abril 4				
Mayo 5				
Junio 6				
Julio 7				
Agosto 8				
Septiembre 9				
Octubre 10				
Noviembre 11				
Diciembre 12				
	$\sum_1^{12} = 313,83$		$\sum_1^{12} = 1721,68$	
	$\frac{\sum_1^{n} (x_i - \bar{x})^2}{n} = 26,15$		$\frac{\sum_1^{n} (x_i - \bar{x})^2}{n} = 143,47$	

2. Respondan entre todos:

- ¿Cuál es el valor de **n** en la fórmula de la varianza?
- ¿Qué indica el subíndice **i**?
- ¿Cuál de las dos ciudades presenta mayor varianza? ¿Por qué?

A

4. Desviación típica

En esta actividad trabajarás sobre el uso de estas medidas de dispersión y sus símbolos a partir de otro problema.

a) Trabajá con la siguiente situación problemática, siguiendo las indicaciones de las consignas.

Para efectuar un control de calidad de los paquetes de arroz que envasa una empresa, se tomaron como muestra 200 paquetes. En la tabla siguiente se muestran los pesos encontrados y la tabla de frecuencias.

Peso en gramos (x_i)	**Frecuencia (f_i)**
498	8
499	60
500	102
501	24
502	6
Total	200

Para calcular la varianza estadística como medida de la dispersión de esta distribución, es necesario calcular primero el promedio de los pesos registrados y las distancias de cada valor a ese promedio.

1. Realizá en tu carpeta una tabla como la siguiente para volcar los resultados que obtengas siguiendo estas indicaciones:

Peso en gramos	Frecuencia	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2\ f_i$
498	8	498 – 499,8 = - 1,8	$-1{,}8^2 = 3{,}24$	3,24 . 8 = 25,92
499	60			
500	102			
501	24			
502	6			
Total	200			

2. Usá la calculadora para realizar las siguientes operaciones:
- Determiná el promedio $\bar{x}$ de los pesos de los 200 paquetes.
- Calculá la distancia de cada valor x_i al promedio.
- Elevá al cuadrado esas diferencias.
- Multiplicá esos cuadrados por las frecuencias respectivas.
- Sumá los valores de la columna $(x_i - \bar{x})^2\ f_i$ y habrás obtenido la varianza de esa distribución.

Si se calcula la raíz cuadrada de la varianza de una distribución de datos, se obtiene lo que en Estadística se llama **desviación típica**.

La desviación típica de una distribución se representa con la letra griega σ (es la letra griega sigma minúscula) e indica la distancia entre los datos de una distribución y el valor del promedio.

En símbolos: $\sigma = \sqrt{\dfrac{\sum_1^n (x_i - x)^2}{n}}$.

La fórmula anterior permite calcular la desviación típica. Su importancia práctica es que cuanto mayor es la desviación típica, más dispersos están los datos con respecto al promedio.

b) A continuación trabajarás con un ejemplo numérico para que te quede más clara la importancia del uso de la desviación típica.

En las tablas siguientes se dan las notas obtenidas por dos grupos de alumnos en un examen.

Grupo A										
Notas	1	2	3	4	5	6	7	8	9	10
Frecuencia	1	0	0	4	10	8	5	6	3	3

Grupo B										
Notas	1	2	3	4	5	6	7	8	9	10
Frecuencia	1	1	2	4	5	5	5	4	3	2

1. Calculá la media de las notas para cada grupo, redondeando los décimos.
2. ¿En cuál de los dos grupos te parece que será mayor la desviación típica?
3. Para comprobar si tu estimación fue correcta calculá la desviación típica de la distribución de datos de cada grupo siguiendo los pasos indicados en el punto **1** de la actividad **4**. Compará tu trabajo con el de tus compañeros.

Habrás podido comprobar que cuanto mayor es la desviación típica, más dispersos están los datos con relación a la media o promedio.

c) Revisá el desarrollo completo de esta unidad y confeccioná con tus compañeros un afiche “para recordar” con el significado de los términos estadísticos que aprendiste: histograma, media o promedio, rango o recorrido, varianza, desviación típica de una distribución. Consulten con el docente en qué libros de la biblioteca pueden encontrar otros problemas interesantes para seguir aprendiendo sobre este tema.

Para finalizar

En esta unidad aplicaste muchos de los conocimientos de Estadística que adquiriste en años anteriores. Revisaste la noción de población o universo sobre el que se realiza una estadística y la conveniencia de trabajar sobre una muestra representativa de toda la población. Organizaste datos en intervalos para construir histogramas que tienen la ventaja de que permiten visualizar una presentación de la totalidad de los datos y sus respectivas frecuencias.

Al revisar las características de los valores centrales o medidas de centralización, viste que si bien el promedio es el más usado, no siempre es representativo de la muestra en estudio. Cuando los valores observados se encuentran muy alejados del promedio, es necesario completar la información con las medidas de dispersión: el rango y la desviación típica son los más útiles.

También aprendiste a usar algunas letras griegas como nuevos símbolos que en Matemática representan elementos característicos de la Estadística. Estos conocimientos te resultarán de utilidad al aplicarlos al estudio de otras áreas de conocimiento como las Ciencias Sociales y las Ciencias Naturales.

DESAFÍOS MATEMÁTICOS

1. Adivina adivinador

Pedile a un amigo que escriba un número de dos cifras sin que vos lo veas, que lo multiplique por 10 y del resultado reste un múltiplo de 9 menor o igual que 81. Pedile que te diga el resultado y vos adivinarás el número que escribió.
Si el resultado es de tres cifras, formá un número con las dos primeras y sumále la cifra de las unidades. En cambio, si el resultado es de dos cifras sumálas entre sí. En ambos casos obtendrás el número secreto. El desafío consiste en que encuentres por qué este procedimiento no puede fallar.

2. Una rollo de tela

Si mido un rollo de tela de 2 metros en 2 metros me sobra uno, si mido de 3 en 3 me sobran dos metros, si mido de 4 en 4 me sobran tres, si lo hago de 5 en 5 me sobran cuatro y de 6 en 6 me sobran cinco. Sabiendo que el rollo tiene menos de 100 metros, ¿cuál es su longitud?

3. Otro sudoku

En la unidad **15** del CUADERNO DE ESTUDIOS **1** te iniciaste en este juego, es muy simple: hay una cuadrícula de 81 cuadrados, organizados en 9 cuadros de 3 × 3, vale decir de 9 cuadrados cada uno. Algunos de estos cuadrados, casi siempre alrededor de 30, ya vienen con una cifra escrita.
Este rompecabezas numérico ideado por Howard Games se publicó por primera vez en Nueva York en 1979 con el nombre de ***Number place*** (el lugar de los números). En 1984 la idea fue introducida en un periódico japonés con el nombre de *Suji wa dokushin ni kagiru* ("los números deben estar solos"), y posteriormente se abrevió esta nomenclatura al nombre por el que hoy se lo conoce en casi todo el mundo: sudoku (números solos).
El objetivo del juego es colocar en los cuadrados vacíos los números que faltan de modo que en cada cuadro de 3 × 3 estén todos los números del 1 al 9, con la condición de que cada número aparezca solamente una vez en cada fila horizontal y en cada columna vertical del cuadro completo.
Resolvé el que aparece a continuación:

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

4. Otro sudoku

Por si te quedaron ganas de resolver otro sudoku.

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