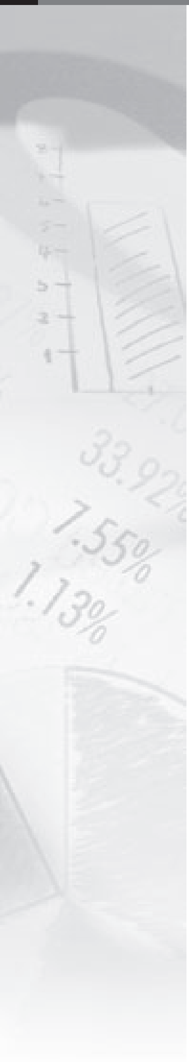
Capítulo IV

# Medidas o estadísticos de resumen




## *Introducción*

Como vimos en el capítulo anterior, un conjunto de datos puede ser representado a través de tablas de frecuencias y gráficos. La principal característica de estas representaciones es que, además de entregar información exacta o cuantitativa, como frecuencias absolutas, o relativas simples o porcentuales, nos entregan información visual de tipo cualitativa. A modo de ejemplo, en un histograma podemos visualizar intervalos de valores donde se encuentran la mayor parte de las observaciones o, en particular, notar que las observaciones están concentradas en valores altos de la variable, habiendo pocas observaciones de valores pequeños, o viceversa. Podemos, también, comparar los histogramas de dos conjuntos de datos diferentes y notar, por ejemplo, que sus formas son similares, pero que se mueven en intervalos diferentes de valores, entre otros aspectos.

Notamos que en afirmaciones como las anteriores nos referimos a características específicas sobre la distribución de los datos, como intervalos de valores más frecuentes o localización de la distribución. Algunas de estas características pueden ser cuantificadas a través de las llamadas *medidas* o *estadísticos de resumen*. En particular, estudiaremos medidas, o estadísticos, de tendencia central, de posición relativa y de dispersión.

Este capítulo está organizado como sigue: la **Sección 1** muestra la importancia de entender e interpretar correctamente las medidas de resumen de distribuciones de datos. Las **Secciones 2**, **3** y **4** se enfocan, cada una, en estadísticos que representan aspectos diferentes de un conjunto de observaciones. En particular, la **Sección 2** trata sobre medidas de tendencia central y se estudian los conceptos de media, mediana y moda. La **Sección 3** se concentra en medidas de posición relativa estudiando, en particular, cuartiles, quintiles, deciles y percentiles, y finalizando con el diagrama de cajón con bigotes, o *boxplot*, que integra estos estadísticos. Finalmente, la **Sección 4** estudia medidas de dispersión o variabilidad de los datos.

# 1. Motivación

Muy frecuentemente encontramos estudios o reportes que describen conjuntos de datos de manera muy sucinta, y entregan, entre otras características, algunos promedios, medianas o desviaciones estándar. Si bien el ciudadano medio posee alguna idea de lo que estos valores representan, resulta importante comprender exactamente su significado, de modo de extraer conclusiones adecuadas, que tomen en cuenta las limitaciones que estas medidas puedan tener. Por otra parte, muchos profesionales deben comunicar hallazgos o resultados de sus estudios o mediciones, para lo cual es necesario que entiendan qué medidas son relevantes en cada caso, sepan la manera de obtenerlas y sean capaces de elaborar conclusiones en torno a los valores obtenidos.

Consideremos, a modo de ejemplo, los resultados del SIMCE, que suelen reportarse a través de una tabla como la **Tabla IV.1**, que muestra los resultados a nivel nacional de las pruebas de Lenguaje, Matemática y Ciencias Naturales de los cuartos básicos, en el año 2011.

| Prueba | Puntaje promedio 2011 |
|---|---|
| Lenguaje | 267 |
| Matemática | 259 |
| Ciencias Naturales | 259 |

*Tabla IV.1: Resultados de las pruebas SIMCE de Lenguaje, Matemática y Ciencias Naturales a nivel nacional, de los cuartos básicos, en el año 2011.*

En la tabla, se leen los promedios de los puntajes de todos los niños del país en cada una de las evaluaciones. Leemos, por ejemplo, que el promedio nacional en la prueba de Lenguaje fue de **267** puntos.

Sin embargo, únicamente a partir de la tabla no podemos responder preguntas como:

- ¿Se puede afirmar que hay igual cantidad de puntajes sobre y bajo el promedio de los resultados nacionales?
- ¿Qué tan cercanos son, en general, los puntajes de cada prueba a su promedio nacional?
- ¿Son los promedios nacionales de los puntajes, buenos representantes de los puntajes de todos los niños del país que rindieron la prueba? ¿en qué sentido?
- Los promedios nacionales en las pruebas de Matemática y Ciencias Naturales son los mismos. ¿Significa esto que los niños dominan sus contenidos en igual medida?

En este capítulo, entregaremos herramientas que nos ayudarán a responder preguntas como las anteriores.

## 2. Medidas de tendencia central

Consideremos un grupo de niños que juega cada uno con un número diferente de cubos, como se muestra en la **Figura IV.1**. Si registramos el número de cubos con que juega cada niño obtenemos:

$$4 \qquad 1 \qquad 5 \qquad 5 \qquad 0$$

*Figura IV.1: Cada niño juega con un número diferente de cubos.*

¿Existe algún número de cubos que sea representativo del número de cubos con que juegan los niños? Aunque en este ejemplo se cuenta con muy pocos datos, sabemos que en el caso de variables cuantitativas podemos representar estos números tanto en una tabla de frecuencias, como en un histograma. A partir de dichas representaciones, sería posible determinar, por ejemplo, que el número de cubos más frecuente es 5. Sin embargo, existen también otros conceptos de *valor representativo*. En efecto, podemos visualizar, al menos, 3 posibles respuestas:

- 5 cubos es el número de cubos más frecuente.
- El promedio de cubos de los niños es 3.
- Si ordenamos las observaciones de menor a mayor, el número de cubos que queda en el centro es 4.

Las tres afirmaciones entregan una idea de valores representativos del número de cubos con que juegan los 5 niños. ¿Cuál de ellas debemos considerar? Las características que se han utilizado para responder a la pregunta se denominan *medidas de tendencia central*. A continuación estudiaremos cada una de ellas: la media o promedio, la mediana y la moda, y notaremos que cada una de ellas tiene una interpretación diferente, por lo que es importante saber lo que estamos comunicando al reportar sus valores.

## 2.1. La media o promedio

En el ejemplo de la **Figura IV.1**, una manera de encontrar un valor de resumen es a través de lo que se denomina *reparto equitativo*. Esto es: si juntamos todos los cubos de los niños y los repartimos entre ellos, asegurándonos de que cada niño reciba la misma cantidad de cubos, la media o promedio corresponde a la cantidad de cubos que recibiría cada niño.

En la **Figura IV.1** vemos que, en total, los niños tienen **15** cubos, pero que no todos juegan con el mismo número de estos. En la **Figura IV.2**, se han repartido de manera equitativa estos **15** cubos, de modo que todos los niños juegan ahora con **3** cubos. Así, decimos que la media o promedio de los cubos de los niños es **3** cubos.

*Figura IV.2: Los cubos han sido repartidos de manera equitativa, y ahora cada uno de los niños tiene 3 cubos. Luego, la media o promedio de la cantidad de cubos es 3.*

Esta idea sugiere una forma de calcular la media: para repartir los cubos, primero podemos juntarlos para saber cuántos cubos debemos repartir. Aritméticamente, esto significa que debemos sumar los números de cubos que tienen los niños. En este caso, obtenemos un total de $4 + 1 + 5 + 5 + 0 = 15$ cubos. Una vez que sabemos cuántos cubos se tienen en total, debemos repartirlos de tal manera que cada uno de los 5 niños reciba la misma cantidad de cubos. Si bien existen diferentes estrategias para repartir los cubos, como, por ejemplo, repartirlos uno a uno siguiendo algún orden en los niños de manera consecutiva, hasta que se nos acaben los cubos, para entender el concepto de media utilizaremos la estrategia de calcular primero cuántos cubos corresponden a cada niño y luego repartir esta cantidad a cada uno, de una sola vez. Esto significa que debemos dividir la suma obtenida por el total de niños, es decir, realizar la operación $\frac{15}{5} = 3$ cubos.

En resumen, el cálculo realizado corresponde a:

$$\frac{4+1+5+5+0}{5} = 3$$

En el caso general en que disponemos de un conjunto de $n$ observaciones, que anotamos como $x_1, x_2, \dots, x_n$, podemos obtener la media de la distribución de las observaciones, que se representa por el símbolo $\bar{x}$, como:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

En el ejemplo que planteamos sobre la cantidad de cubos que tiene cada niño de la **Figura IV.1**, partiendo con el niño de arriba a la izquierda, $x_1 = 4$, $x_2 = 1$, $x_3 = 5$, $x_4 = 5$, $x_5 = 0$ y $n = 5$, y la fórmula recién descrita corresponde a

$$\bar{x} = \frac{4+1+5+5+0}{5} = 3$$

Esta operación ya la habíamos realizado con anterioridad. Notemos que dado que la suma es conmutativa, el orden en que recolectamos las observaciones no es relevante.

Dijimos anteriormente que, al pensar en repartir los cubos a los niños, no estamos pensando en la repartición de los cubos uno a uno. En efecto, supongamos que, en total, los niños tuviesen 12 cubos. Si los repartimos uno a uno siguiendo un orden preestablecido de los niños, al entregar el décimo cubo habremos entregado un total de 2 cubos a cada uno de los 5 niños. Al repartir los 2 cubos restantes a los 2 niños siguientes según el ordenamiento, tendremos 2 niños con 3 cubos y 3 niños que tendrán solo 2 cubos. No habremos logrado el objetivo del reparto equitativo.

El ejemplo recién mencionado muestra también que, si bien la interpretación de la media a través del reparto equitativo es una de las más intuitivas, presenta ciertas limitaciones. En efecto, en el caso anterior, la media del número de cubos corresponde a $\frac{12}{5} = 2{,}4$ cubos. ¿Cómo pueden repartirse 2,4 cubos a cada niño, si no podemos particionar los cubos? Esta es una de las características de la media: ella puede tomar valores que no corresponden a valores válidos de la variable, lo cual genera problemas en una representación a través del concepto de reparto equitativo, cuando los objetos no son particionables, como es el caso de los cubos. En situaciones como esta, no es aconsejable utilizar la interpretación de reparto equitativo para la media.

Otra manera de interpretar la media es a través del concepto de nivelación de los valores de las observaciones. A modo de ejemplo, consideremos 3 varillas de alturas 30, 20 y 40 centímetros, y supongamos que las ubicamos una al lado de la otra, como se muestra en la **Figura IV.3**, arriba. Como las 3 varillas tienen diferentes alturas, no sería posible apoyar un objeto sobre ellas de manera perfectamente horizontal. Si cortamos ahora la varilla de 40 centímetros en dos trozos de 30 y 10 centímetros, respectivamente, como muestra la **Figura IV.3**, al centro, y añadimos el trozo de 10 centímetros a la varilla de 20, como se muestra en la **Figura IV.3**, abajo, las 3 varillas resultantes tendrán una altura de 30 centímetros, que corresponde al valor de la media, y se podría apoyar un objeto sobre ellas de manera perfectamente horizontal. Esta idea puede visualizarse, por ejemplo, al considerar la construcción de una mesa. Si las patas a utilizar tienen diferentes alturas, la cubierta de la mesa quedará inclinada. Si pudiéramos realizar cortes y uniones a las patas, el procedimiento de nivelación descrito lograría que la cubierta de la mesa quede perfectamente horizontal y la altura de cada pata correspondería al promedio de sus alturas originales.

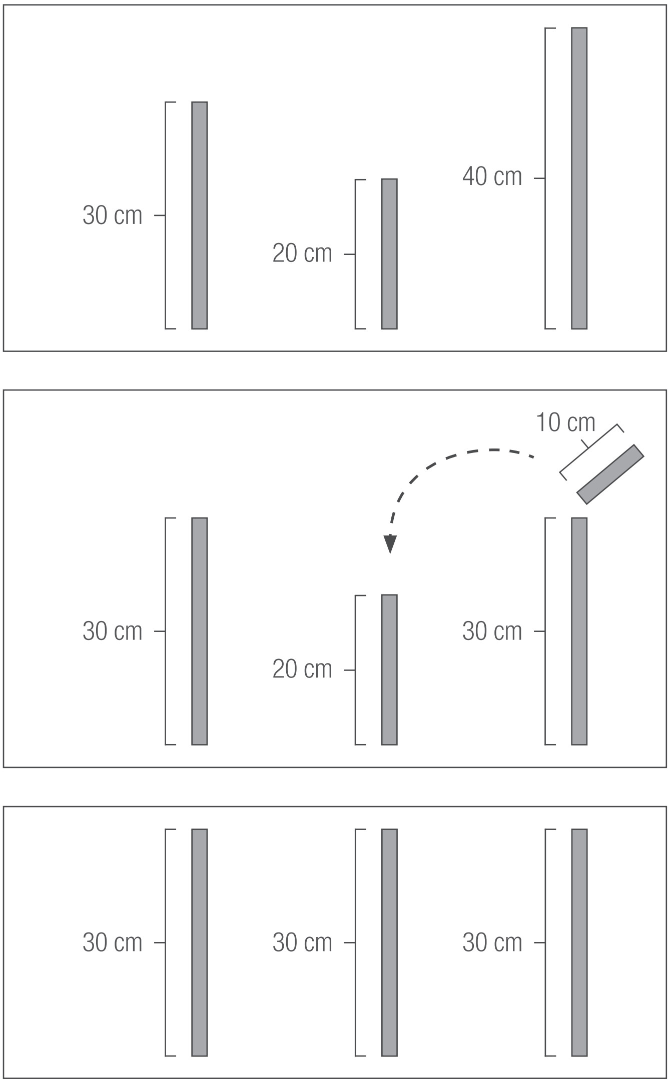



*Figura IV.3: La media como nivelación. Arriba: las varillas tienen diferentes alturas. Centro: quitamos un trozo de 10 cm a la varilla más alta. Abajo: unimos este trozo a la varilla más baja. Las tres varillas resultantes tienen ahora la misma altura, 30 cm. Esta altura común corresponde a la media de las alturas de las barras.*

Otra manera de interpretar la media corresponde a un concepto de igualdad de sumas de distancias, idea que ilustraremos a través de un ejemplo. Consideremos las notas de 4 niños de un curso:

3,0 7,0 7,0 7,0

Al realizar las operaciones, encontramos que la media de este conjunto de datos corresponde a un **6,0**. Veremos que esta nota es la única nota tal que la suma de las distancias desde sí misma a las observaciones que son menores a ella es igual a la suma de las distancias desde sí misma a las observaciones que son mayores.

Para ayudar a la comprensión, comenzaremos considerando solo 2 niños: el niño que obtuvo la nota **3,0**, y uno de los 3 niños que obtuvo nota **7,0**. La media de las notas de estos 2 niños es **5,0**. La **Figura IV.4** muestra las 2 observaciones consideradas, a través de círculos en los valores **3** y **7**, y sus distancias a la nota **5,0**. La distancia desde la observación menor que **5,0** y este último es $|3 - 5| = 2$, y la distancia desde la observación que es mayor que **5,0** y este último es $|7 - 5| = 2$, logrando que las distancias a ambos lados de la media, **5,0**, sean las mismas, como habíamos adelantado.

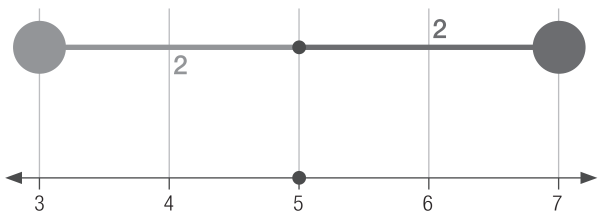



*Figura IV.4: La distancia entre la notas 3,0 y 5,0 es $|3 - 5| = 2$, la misma que la distancia entre las notas 7,0 y 5,0, $|7 - 5| = 2$.*

Incorporemos ahora a los 2 niños restantes que obtuvieron un **7,0**. La **Figura IV.5** muestra las 4 observaciones mediante círculos, y sus distancias a la nota **5,0**. La distancia desde la nota menor que **5,0** y esta última es:

$$|3 - 5| = 2$$

Mientras que las distancias desde las notas mayores que **5,0** y esta última es:

$$|7 - 5| = 2 \qquad |7 - 5| = 2 \qquad |7 - 5| = 2$$

Cuya suma es **6**. Luego, la suma de las distancias de las observaciones menores que **5,0**, que es **2**, no es igual a la suma de las distancias de las observaciones que son mayores, que es **6**. Luego, la nota **5,0** ya no cumple con la condición pedida.

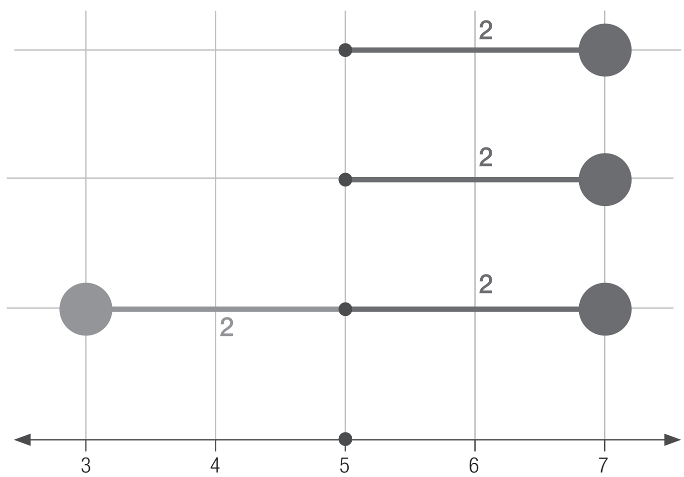



*Figura IV.5: La distancia desde la nota menor a 5,0 es $|3 - 5| = 2$. Las distancias desde las tres notas mayores que 5,0 son todas iguales a $|7 - 5| = 2$, por lo que su suma, 6, es diferente a la anterior.*

Dado que la suma de las distancias de las observaciones mayores que la nota 5,0 es mayor, podemos intuir que debemos mover el punto buscado hacia la derecha, con el objeto de disminuir las distancias desde este a las observaciones mayores. Veremos que esto realmente ocurre. En efecto, notemos que, algebraicamente, lo que se quiere es encontrar un punto *a* tal que:

$$|3 - a| = |7 - a| + |7 - a| + |7 - a|$$

Asumiendo que el punto buscado es mayor que 3,0 y menor que 7,0, podemos eliminar los valores absolutos en la expresión de la manera:

$$a - 3 = (7 - a) + (7 - a) + (7 - a)$$

Es decir, $-3 - 7 - 7 - 7 + 4a = 0$. Despejando el valor de *a*, obtenemos que este debe ser igual a:

$$a = \frac{3 + 7 + 7 + 7}{4} = 6$$

lo que coincide con la media de las notas de los 4 niños.

Esta situación se ilustra en la Figura IV.6, que muestra que la distancia desde la observación menor que la media, que es 3, es igual a la suma de las distancias desde las observaciones mayores que la media, que también es $(1 + 1 + 1) = 3$.

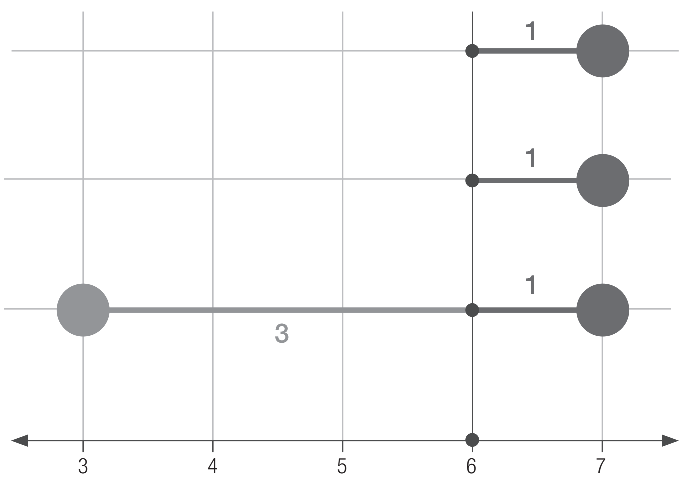



*Figura IV.6: La distancia entre la nota menor que la media y esta última, que es 3, es igual a la suma de las distancias entre las 3 notas mayores que la media y esta última, que es (1+1+1 = 3).*

Notemos, que en el ejemplo que seguimos, solo existen 2 valores en el conjunto de datos, 3,0 y 7,0. Se puede mostrar que la condición de igualdad de sumas de distancias se cumple en cualquier conjunto de observaciones, aunque existan más de dos valores diferentes en él. El siguiente ejemplo ilustra un caso como este.

**Ejemplo**

Suponga que las notas de otros 4 alumnos son:

4,0 5,0 5,5 6,3

Cuya media es 5,2. La suma de las distancias desde las observaciones menores a 5,2 a este punto es:

$$|4 - 5{,}2| + |5{,}0 - 5{,}2| = 1{,}4$$

Y la suma de las distancias desde las observaciones mayores a 5,2 a este punto es:

$$|5{,}5 - 5{,}2| + |6{,}3 - 5{,}2| = 1{,}4$$

Lo que es igual a la distancia anterior.

Se puede demostrar que, en todo conjunto de observaciones, la media corresponde al único valor que cumple con la propiedad pedida.

De este modo, en general, podemos interpretar la media como el punto tal que la suma de las distancias desde las observaciones a su izquierda y este mismo es igual a la suma de las distancias desde las observaciones a su derecha y este mismo.

Notemos que, a diferencia de lo que ocurre con la interpretación de reparto equitativo cuando los elementos a repartir no son particionables, tanto la interpretación de la media como nivelación, como la de igualdad de sumas de distancias pueden utilizarse sin necesidad de que la media tome valores enteros. En efecto, las barras verticales pueden nivelarse a cualquier altura, así como también el punto que cumple la condición de igualdad de sumas de distancias puede encontrarse en cualquier ubicación entre los datos.

### *En resumen*

- La *media* o *promedio* corresponde a una medida de tendencia central, que puede ser interpretada en base a los conceptos de reparto equitativo, nivelación e igualdad de sumas de distancias.
- El concepto de *reparto equitativo* sugiere la manera de obtener el valor de la media: "juntamos" (sumamos) los valores de las observaciones como si fueran unidades, y las repartimos equitativamente (dividimos) entre todas ellas.
- La media de un conjunto de datos puede no ser un valor admisible de la variable de interés, situación que dificulta su interpretación a través del concepto de reparto equitativo.

Notemos que, muchas veces, se presenta la media a través de las operaciones aritméticas necesarias para obtener su valor. A modo de ejemplo, suele encontrarse que "la media corresponde a la suma de los valores de las observaciones, dividida por el número total de ellas". Sin embargo, el énfasis debe estar en comprender lo que el concepto de media representa. Hemos mostrado aquí

3 interpretaciones importantes: reparto equitativo, nivelación e igualdad de sumas de distancias, que explican de qué forma la media es un valor representativo de los valores de las observaciones.

Notemos que solo tiene sentido hablar de la media, o promedio, cuando los datos son cuantitativos. Al igual como se hizo al describir variables cualitativas en el **Capítulo 3**, hacemos hincapié en que no se deben confundir las variables cualitativas que han sido codificadas con etiquetas numéricas como, por ejemplo, rojo = 1, azul = 2, etc., con variables cuantitativas. Esto también es válido para las variables cualitativas ordinales, como grados de acuerdo y desacuerdo. En efecto, aunque por tener categorías ordenables podríamos asignar números a sus valores, no existe el concepto de distancia entre categorías que determine las distancias entre los valores asignados.

## 2.2. La mediana

Supongamos que nos interesa estudiar la altura de los niños del quinto básico de una escuela. Podemos encontrar un concepto de centro de las alturas ubicando a los niños en una fila en orden creciente (o decreciente) de altura y encontrando al niño que queda en la posición central. Notamos que la altura de dicho niño corresponde a una característica del centro de las alturas de los niños. De este modo, llamamos *mediana* de la altura de los niños del curso, a la altura de la niña ubicado en la posición central, como se ilustra en la **Figura IV.7**.


*Figura IV.7: Los niños han sido ordenados desde el más pequeño al más grande. La mediana corresponde a la altura de la niña central.*

En otro ejemplo, consideremos las notas finales obtenidas por un grupo de 13 alumnos en un curso, las que han sido ordenadas de manera creciente:

4,3 4,6 4,6 4,7 5,0 5,0 **5,0** 5,0

5,7 5,9 6,1 6,1 6,6

La observación central corresponde al puntaje en negrita, puesto que existen 6 observaciones a su izquierda y 6 observaciones a su derecha. Luego, la mediana de las notas de este curso corresponde a 5,0. Notemos que se obtiene el mismo resultado al ordenar las observaciones en orden creciente o decreciente, lo que ocurre con la mediana en todo conjunto de datos.

Tanto en el problema de las alturas de los niños, en la **Figura IV.7**, como en el de las notas finales de un curso, el número de observaciones en el conjunto de datos ha sido impar, **5** y **13**, respectivamente. En este caso, al ordenar los valores de menor a mayor, es posible determinar la observación que está al centro dejando igual número de observaciones a su izquierda y a su derecha. Sin embargo, esto no es posible si el número de observaciones es par.

En efecto, consideremos nuevamente las edades, medidas en años, de los **34** profesores de una escuela, que tratamos en el **Capítulo 3**. Las edades han sido ordenadas de manera creciente y quedaron como sigue:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 31 | 32 | 32 | 32 | 33 | 35 | 36 | 37 | 37 | 38 |
| 39 | 39 | 40 | 40 | 41 | 42 | **42** | **43** | 43 | 44 |
| 45 | 46 | 47 | 48 | 48 | 51 | 53 | 55 | 56 | 56 |
| | | | 60 | 61 | 62 | 76 | | | |

En este caso, no existe una única observación que esté al centro. En efecto, ambas observaciones en negrita, **42** y **43** años, corresponden a observaciones centrales, en el sentido de que hay **16** observaciones a la izquierda de **42**, y **16** observaciones a la derecha de **43**. Debemos extender el concepto de mediana, para lo que diremos que corresponde a un valor, no necesariamente una observación en el conjunto de datos, que deja el mismo número de observaciones a su izquierda que a su derecha. El valor **42** años no cumple con la condición, puesto que deja a su izquierda **16** observaciones y, a su derecha **17**, dos cantidades diferentes. Algo similar ocurre con la observación **43** años.

Según el concepto anterior, la mediana deja de ser única. Efectivamente, en el ejemplo, cualquiera de los infinitos valores entre **42** y **43** años cumple con esta propiedad. Si tomamos como ejemplo el valor **42,7** años, este deja el mismo número de observaciones a su izquierda, que es **17**, y a su derecha. En estos casos, lo habitual es convenir que la mediana corresponda al promedio entre las observaciones centrales. En el ejemplo se obtiene:

$$\text{Mediana} = \frac{42 + 43}{2} = 42{,}5 \text{ años}$$

Notamos que, al igual que con la media, la mediana puede no corresponder a un valor posible de la variable de interés.

El diagrama en la **Figura IV.8** resume el procedimiento para encontrar la mediana, tanto cuando el número de observaciones es par, como impar.

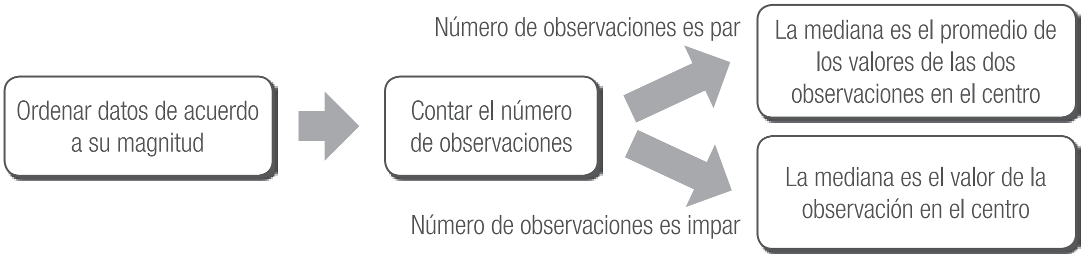



*Figura IV.8: Esquema del procedimiento para obtener el valor de la mediana de un conjunto de observaciones.*

Hasta aquí hemos introducido la mediana a través de la idea de *valor central*. Al igual que con la media, veremos también otras interpretaciones de la mediana. En particular: de la idea de valor central, se desprende que *la mediana es un punto que divide a las observaciones en dos grupos de igual tamaño*, y por otra parte, *la mediana es tal que, al menos, el 50% de las observaciones es menor o igual a ella y, al menos, el 50% de las observaciones es mayor o igual que la misma*.

Derivaremos ambas interpretaciones a través de un ejemplo. Para la primera de ellas, consideremos nuevamente el problema de estudiar la cantidad de cubos que tiene cada uno de los 5 niños en la **Figura IV.1**. Si ordenamos los números de cubos de menor a mayor, obtenemos:

$$\overbrace{0 \quad 1}^{\text{Dos observaciones}} \quad 4 \quad \overbrace{5 \quad 5}^{\text{Dos observaciones}}$$

Mediana = 4

Donde la mediana corresponde al dato central, 4 cubos. Vemos que la mediana divide a las observaciones en 2 grupos, cada uno con 2 observaciones.

Consideremos también los siguientes datos, que corresponden a las edades de las madres de 12 niños en un Taller de Guitarra:

$$\overbrace{33 \quad 34 \quad 35 \quad 37 \quad 39 \quad 39}^{\text{Seis observaciones}} \quad \overbrace{39 \quad 41 \quad 42 \quad 43 \quad 45 \quad 50}^{\text{Seis observaciones}}$$

Mediana = 39

Según la convención que tomamos, la mediana en este caso corresponde al promedio de las dos observaciones centrales. Por ser ambas iguales a 39 años, su promedio corresponde también a 39 años. Vemos que la mediana dividió a las observaciones en dos grupos, a su izquierda y a su derecha, y que nuevamente, estos contienen el mismo número de observaciones que, en este caso, es 6.

Esto nos da entonces la interpretación de la mediana como un punto que divide al conjunto de observaciones, al ordenarlas en orden creciente (o decreciente) de magnitud, en dos grupos de igual tamaño. Notemos que esto ocurre tanto cuando el número de observaciones es par como impar, según vimos en los dos ejemplos recientes.

Para la segunda interpretación, *la mediana es tal que, al menos, el 50% de las observaciones es menor o igual a ella y, al menos, el 50% de las observaciones es mayor o igual ella*, consideremos nuevamente el número de cubos de los 5 niños:

$$0 \quad 1 \quad 4 \quad 5 \quad 5$$

Donde la mediana es 4 cubos. En este caso, las observaciones menores o iguales a la mediana son 3: 0 cubos, 1 cubo y 4 cubos. Es decir, $\frac{3}{5}$ de las observaciones, o un 60%, son menores o iguales a la mediana. Por otra parte, las observaciones mayores o iguales a la mediana también son 3: 4 cubos, 5 cubos y 5 cubos. Es decir, un 60% de las observaciones es mayor o igual a la mediana.

Vemos que en este conjunto de datos la mediana cumple con la propiedad deseada, dado que ambos porcentajes son mayores a 50%.

En el ejemplo anterior, ocurre que los porcentajes referidos son iguales, lo cual no se cumple en todo conjunto de observaciones. En efecto, consideremos nuevamente las edades de las madres:

33 34 35 37 39 39 | 39 41 42 43 45 50

Mediana = 39

Vimos que la mediana corresponde a 39 años. En este caso, el número de observaciones menores o iguales a la mediana es 7, lo cual corresponde a $\frac{7}{12}$ de las observaciones o, aproximadamente, 58%. Por otra parte, el número de observaciones mayores o iguales a la mediana es 8, lo cual corresponde a $\frac{8}{12}$ de las observaciones o, aproximadamente, 67%. Vemos que, en este caso, la mediana cumple con las condiciones pedidas, dado que ambos porcentajes son mayores a 50%; sin embargo, los porcentajes involucrados no son los mismos. Esto se debe a que el conjunto de datos contiene más de una observación que toma el valor de la mediana, y a que los 2 grupos formados contienen diferente número de estas repeticiones: 2 y 1 valores iguales a 39 en cada grupo. Este es el caso general.

Consideremos ahora un tercer ejemplo, que se refiere a los puntajes SIMCE de Lenguaje de cierto nivel y año, de todos los alumnos del país.

*Para pensar*

Si se nos indica que "la mediana de todos los puntajes de los niños es 260 puntos". ¿Tiene importancia, en este caso, que distingamos si los porcentajes de observaciones menores o iguales a 260 puntos y de observaciones mayores o iguales a 260 puntos son exactamente iguales? Justifique su respuesta.

En este ejemplo, el número de alumnos que rinde la prueba SIMCE cada año es considerablemente grande y es esperable que el número de observaciones iguales a la mediana sea relativamente pequeño, en comparación al número total de estos alumnos. Distinguir, en este caso, diferencias en los porcentajes carece de sentido, y podemos decir que "aproximadamente, un 50% de las observaciones es menor o igual que la mediana", y que "aproximadamente, un 50% de las observaciones es mayor o igual que la mediana".

## En resumen

- La *mediana* corresponde a un valor que queda al centro al ordenar las observaciones creciente o decrecientemente. Si el número de observaciones es par, no existe un único valor "al centro" y, en ese caso, se propone elegir la mediana como el promedio de los valores de las dos observaciones centrales.
- La mediana también puede interpretarse como un valor que divide al conjunto de observaciones, ordenado según su magnitud, en 2 grupos de igual tamaño.
- La mediana también puede interpretarse como un valor tal que, al menos un 50% de las observaciones son menores o iguales a ella y, al menos, un 50% de las observaciones son mayores o iguales a ella.

Notamos que, en general, la mediana solo tiene sentido para observaciones o variables cuantitativas. Si bien es posible ordenar los valores de variables cualitativas ordinales, en general, estas variables presentan muy pocas categorías, por lo que la mediana carece de utilidad y no se recomienda su uso.

## Ejercicios

1. Usted desea explicar a sus alumnos que la tasa de natalidad del país en el año 2011 fue de 1,87 infantes nacidos por mujer. ¿Qué interpretación o representación de la media utilizaría para este fin?

2. Un carpintero dispone de 4 palos de madera que utilizará para construir una mesa. Estos palos miden 100, 82, 106 y 70 cm, respectivamente. Suponiendo que él puede trozar la madera y añadir cortes, represente los palos como en la **Figura IV.3** y muestre una secuencia para lograr que todos los palos tengan la misma altura, de modo que la cubierta de la mesa quede perfectamente horizontal.

3. Los siguientes datos corresponden a las edades, en años, de las madres de los alumnos del sexto básico:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 35 | 34 | 37 | 33 | 38 | 45 | 40 | 41 | 50 | 43 | 39 | 42 |

   Obtenga la media y la mediana de estos datos, y comente la similitud o diferencia entre estos valores.

4. Encuentre la media y la mediana de las alturas (en centímetros) de los 7 alumnos en el Taller de Teatro de un curso:

| | | | | | | |
|---|---|---|---|---|---|---|
| 153,8 | 154,7 | 156,9 | 154,3 | 152,3 | 156,1 | 152,3 |

5. Se ha determinado que un exceso de la concentración de plomo en el aire puede producir efectos nocivos para la salud. En efecto, se ha determinado el valor de 1,5 microgramos por metro cúbico como nivel máximo tolerable. Los siguientes datos corresponden a concentraciones de plomo, en microgramos por metro cúbico, medidos en el aire, al interior de un edificio de una escuela:

5,4 1,1 0,42 0,73 0,48 1,1

a. Obtenga la concentración de plomo media, o promedio, en estas 6 mediciones.

b. Obtenga la mediana de la concentración de plomo en estas 6 mediciones.

c. Sabemos que ambos valores, la media y la mediana, corresponden a valores representativos del conjunto de datos. ¿A qué cree usted que se debe la diferencia obtenida entre la media y mediana de este conjunto?

6. Para ejercitar el concepto de media, una profesora pide a los niños obtener el número promedio de lápices de colores que tiene cada uno de ellos en su estuche. Las siguientes son las cantidades que dice cada uno de los 15 niños del curso:

4 2 1 5 10 3 7 9

2 5 5 2 7 1 12

a. Obtenga la media de este conjunto de datos.

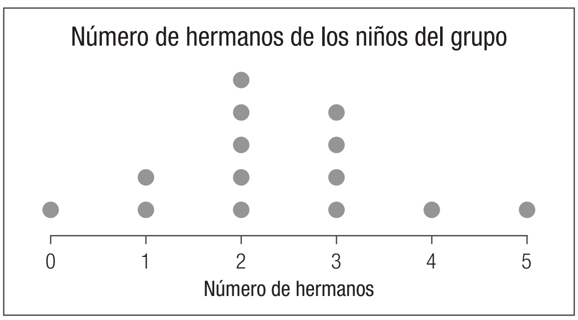
b. ¿De qué manera puede usted interpretar la media en este contexto, para explicarla a sus alumnos?

c. Uno de los 15 niños se ha equivocado al decir que tiene 10 lápices, pues en realidad tiene 20. Obtenga la media reemplazando la observación equivocada por la correcta, de 20 lápices. ¿Cómo cambia la media al compararla con su respuesta en a.?

d. Obtenga la mediana de los datos originales reportados y luego, la mediana al corregir la observación de valor 10 lápices por 20 lápices. ¿Cómo cambia la mediana?

7. Uno de sus alumnos está interesado en conocer los resultados en la competencia del salto largo de su escuela. Un compañero le cuenta que el atleta que obtuvo el quinto lugar, entre 9 competidores, saltó una distancia de 2,15 m. ¿A qué medida de tendencia central corresponde esta longitud?

8. Considere el siguiente diagrama de puntos sobre el número de hermanos de un grupo de niños:




a. Entregue una estimación de la media de este conjunto de observaciones, sin obtenerla de manera exacta, basándose en la interpretación de igualdad de sumas de distancias.