Medidas de posición

Seguramente, has oído hablar del promedio de una serie de datos o del dato que se repite más veces. Son herramientas que se utilizan en muchos casos de la vida diaria.

En esta lectura estudiaremos los valores más importantes que pueden calcularse en una distribución de datos en una tabla de frecuencias. Por ahora estudiaremos los de posición o de tendencia central. También es importante qué forma le dan esos valores a la distribución. Ellos marcan una tendencia que luego da lugar a un análisis de los datos recogidos. Acompáñanos a descubrir cuáles son y cómo se calculan.

Presentación de la situación práctica

Caso CORLUX

CORLUX es una fábrica de la Provincia de Córdoba que se dedica a la producción de una línea de lámparas de bajo consumo. CORLUX ha estado fabricando lámparas durante más de 30 años, pero el rápido giro que se ha registrado en la fabricación de lámparas hacia las de menor consumo, mayor duración y cuidado del medioambiente ha obligado a la empresa a evolucionar durante los últimos 10 años: ha tenido que adecuarse en materiales, procesos de producción y capacitación del personal, entre otras cuestiones.

Debido a la línea nueva que ha lanzado, CORLUX quiere posicionarse en el mercado por la buena calidad de su nuevo producto. Por esta razón, el Departamento de Control de Calidad ha venido realizando en los últimos años acciones para verificar la calidad de su producto: por un lado, viene tomando, después de cada remesa, muestras aleatorias de cajas de 12 unidades cada una para verificar que todas las lámparas estén en condiciones de salir al mercado. En el siguiente cuadro, se registran los resultados obtenidos de la última muestra tomada:

La tabla registra la cantidad de lámparas defectuosas que se encontraron en una muestra aleatoria realizada por CORLUX.

Tabla 1: Lámparas con algún defecto muestreadas por la empresa CORLUX

Cantidad de lámparas con algún defecto	N.° de cajas
0	10
1	6
2	4
3	4
4	5
5	1

Fuente: elaboración propia.

Por otro lado, midió la vida útil de las lámparas de la nueva línea. Esto lo ha venido haciendo en los últimos años, en un laboratorio de pruebas. El último año ha tomado una muestra aleatoria de 90 lámparas, y los resultados se resumen en la siguiente tabla.

La tabla muestra los registros, de menor a mayor, de las horas que duraron las lámparas muestreadas al realizar un control de calidad.

Tabla 2: Vida útil de las lámparas muestreadas en CORLUX

N.° de clase	Vida útil en horas [Li-Ls)	N.° de lámparas
1	215-515	1
2	515-815	1
3	815-1115	2
4	1115-1415	5
5	1415-1715	15
6	1715-2015	18
7	2015-2315	22
8	2315-2615	25
		90

Fuente: elaboración propia.

El jefe del Departamento de Control de Calidad desea dos informes para tener en cuenta y poder tomar decisiones al respecto.

1. Elaborar un informe con los datos de la Tabla 1 sobre las fallas en las lámparas, con base en los estadísticos de posición.
2. Elaborar un informe con los datos de la Tabla 2 sobre la vida útil de las lámparas, con base en los estadísticos de posición.

Medidas de posición para datos sin agrupar

Bajo este tema nos adentramos en la resolución del primer informe que requiere el jefe de Control de Calidad de CORLUX.

¿Por qué el primer informe corresponde a datos sin agrupar? ¿Cuál es la variable en estudio? ¿Qué tipo de variable es?

Recuerda que la primera columna corresponde a la variable en estudio xi, y la segunda, en el caso del cuadro 1, a la frecuencia absoluta fi. Por lo tanto, la variable en estudio es la cantidad de lámparas defectuosas, que es cuantitativa discreta. Además, toma pocos valores distintos, por lo que no hace falta agruparla en intervalos.

Comencemos entonces a estudiar las herramientas necesarias para resolver la situación de la fábrica CORLUX.

Medidas de posición (o de tendencia central)

¿Qué es una medida de posición?

“Es la medida que describe cómo todos los valores de los datos se agrupan en torno a un valor central” (Berenson, Levine y Krehbiel, 2006, p. 72).

Lo que nos proponemos es estudiar valores característicos de una distribución de datos.

Son medidas que resumen los datos de una distribución.

¿Cuáles son las medidas de posición más importantes?

Media aritmética o promedio

Es el cociente entre las sumas de todas las observaciones y el número total de ellas.

- A la media poblacional la representamos con la letra μ y al tamaño de la población con *N*, por lo que la media poblacional se expresa así:

$$\mu = \frac{x_1 + x_2 + \cdots + x_N}{N}$$

Y en notación abreviada:

$$\mu = \frac{\sum_1^N x_i}{N}$$

- En cambio, a la media muestral la representamos con $\bar{\mathbf{x}}$ y al tamaño de la muestra o número de observaciones lo representaremos con **n**.

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

Y en notación abreviada:

$$\bar{x} = \frac{\sum_1^n x_i}{n}$$

Ejemplo 1
En un brote de gripe invernal, 8 personas que desarrollaron la enfermedad tardaron la siguiente cantidad de días en curarse completamente:8, 8, 9, 10, 7, 7, 7, 9. Para calcular la media de esta población, procedemos de la siguiente manera:

$$\mu = \frac{8+8+9+10+7+7+7+9}{8}$$

$$\mu = \frac{65}{8} \rightarrow \mu = 8{,}125$$

Es viable notar que, si bien la media no coincide con ninguna medición, nos representa el valor medio de duración de la enfermedad.

Interpretación de la media:

En promedio, el brote invernal duró 8,125 días en los pacientes.

Mediana

Es el valor que divide en partes iguales a los datos ordenados de una distribución. La mediana, en una distribución ordenada, deja a uno de sus lados los valores menores o iguales a ella y hacia el otro lado los mayores o iguales a ella, en igual número de observaciones.

Para calcularla, se procede de las siguientes dos maneras, según si el número de observaciones fuera par o impar:

1. Si *n* es impar, es el valor central al ordenar la distribución, por ejemplo:

Tabla 3: Ejemplo de un arreglo ordenado de datos correspondiente a las edades de una muestra de 5 personas, es decir, n es impar

Arreglo ordenado	22	23	25	29	30
Orden de la distribución	1	2	3	4	5

$\tilde{x} = 25$

Fuente: elaboración propia.

Observa que la mediana no es 3, sino el valor que ocupa el lugar 3 del arreglo ordenado.

Interpretación de la mediana: El 50 % de las personas tiene una edad menor o igual a 25 años, y el otro 50 %, una edad mayor o igual a 25 años.

2. En cambio, si *n* es par, es el promedio de los valores centrales al ordenar la distribución. Por ejemplo:

Tabla 4: Ejemplo de un arreglo ordenado de datos correspondiente a las edades de una muestra de 6 personas, es decir, *n* es par

Arreglo ordenado	32	35	37	40	42	83
Orden de la distribución	1	2	3	4	5	6

Fuente: elaboración propia.

$$\boxed{\tilde{x} = \frac{37+40}{2} = 38{,}5}$$

Interpretación de la mediana: El 50% de las personas tiene una edad menor a 38,5 años, y el otro 50%, una edad mayor a 38,5 años.

La mediana no es sensible a los valores extremos, por lo que —según el caso— describe con más exactitud las distribuciones con valores extremos despegados de la mayoría.

El promedio es evidente que daría mayor porque hay una persona con 83 años. La más representativa, en este caso, es la mediana.

Moda

Es el valor que más se repite en una distribución.

Ejemplo 2

Si las notas de Estadística de 10 alumnos muestreados en un curso son3, 5, 7, 6, 7, 6, 5, 10, 8 y 7:

El valor de la variable que más se repite es 7.
Por lo tanto, la moda es 7.

$$\hat{x} = 7$$

Si hay dos valores de la variable que más se repiten, es decir, se repiten la misma cantidad de veces y esta cantidad es la mayor, la distribución tiene dos modas y se llama bimodal; si las modas son tres, trimodal; o multimodal, si se dieran varias modas.
Entonces: puede existir una o más modas.
En cambio, si todos los valores de la distribución tienen el mismo número de repeticiones, se dice que no tiene moda.

Interpretación de la moda: La nota que más se da es 7.

Antes de avanzar, hagamos un resumen de la notación y las fórmulas de las medidas de posición estudiadas hasta aquí:

MEDIA | MEDIANA | MODA

Tabla 5: Resumen de la notación y las fórmulas de la media, diferenciando población y muestra

	Notación	Fórmula
Población	μ	$\mu = \frac{\sum_1^N x_i}{N}$
Muestra	$\bar{x}$	$\bar{x} = \frac{\sum_1^n x_i}{n}$

MEDIA | MEDIANA | MODA

Tabla 6: Resumen de la notación y las fórmulas de la mediana, diferenciando población y muestra

	Notación	Fórmula
Población	Me	$Me=\begin{cases} dato\left[pos\ \frac{N+1}{2}\right]\ si\ N\ es\ impar \\ \frac{dato\left[pos\left(\frac{N}{2}\right)\right]+dato\left[pos\left(\frac{N}{2}+1\right)\right]}{2}\ si\ N\ es\ par \end{cases}$
Muestra	$\tilde{x}$	$\tilde{x}=\begin{cases} dato\left[pos\ \frac{n+1}{2}\right]\ si\ n\ es\ impar \\ \frac{dato\left[pos\left(\frac{n}{2}\right)\right]+dato\left[pos\left(\frac{n}{2}+1\right)\right]}{2}\ si\ n\ es\ par \end{cases}$

Fuente: elaboración propia

MEDIA | MEDIANA | MODA

Tabla 7: Resumen de la notación de la moda, diferenciando población y muestra

	Notación	Fórmula
Población	Mo	$no\ hay$
Muestra	$\hat{x}$	$no\ hay$

Fuente: elaboración propia

¿Cómo calcular las medidas de posición cuando los datos están en una tabla sin agrupar?

Ahora bien, tratándose de un grupo de datos más grande, donde algunos de ellos se presentan con distintas frecuencias, puede dárseles un tratamiento distinto a los datos para obtener las medidas de posición.

Vamos a calcular la media, la mediana y la moda para el siguiente ejemplo.

Ejemplo 3

En la siguiente tabla, se registraron las edades de una muestra aleatoria de 20 empleados correspondiente a una empresa gastronómica con varias sucursales en la ciudad de Buenos Aires. Se pide calcular las medidas de posición e interpretarlas para el caso.

Se muestran las edades registradas por la empresa de una muestra de 20 empleados.

Tabla 8: Edades de los empleados de una empresa gastronómica

Edad de cada empleado (xi)	Cantidad de empleados (fi)
18	3
19	4
20	4
21	3
22	3
28	2
35	1

Fuente: elaboración propia.

Cálculo de las medidas de posición para el ejemplo 3, Tabla 8

Media

Cálculo de la media: como deben sumarse todos los datos de la distribución, imaginemos cómo sería el arreglo ordenado de la Tabla 6:
18, 18, 18, 19, 19, 19, 19.........................35. Obviamente, resulta más sencillo multiplicar las edades por su respectiva frecuencia. Por lo tanto, en general:

$$\bar{x} = \frac{x_1 \times f_1 + x_2 \times f_2 + \cdots + x_n \times f_n}{f_1 + f_2 + \cdots + f_n}$$

Llamando f_1, f_2 ...f_n a las distintas frecuencias respectivas a los datos, nos queda la siguiente fórmula:

Fórmula de la media:

$$\bar{x} = \frac{\Sigma(x_i . f_i)}{\Sigma f_i}$$

Pues se trata de una muestra.

En el caso del ejemplo 3, Tabla 6:

$$\bar{x} = \frac{18 \times 3 + 19 \times 4 + 20 \times 4 + 21 \times 3 + 22 \times 3 + 28 \times 2 + 35 \times 1}{20}$$

$$\bar{x} = \frac{430}{20} = 21{,}5$$

Interpretación: el promedio de edad de los empleados es de 21,5 años.

Mediana

Cálculo de la mediana: para encontrar el dato que ocupa la posición central en el arreglo ordenado, es más fácil si confeccionamos la columna de frecuencias acumuladas. Por otra parte, sabemos que hay dos términos centrales porque el número de observaciones es par.

¿Qué posiciones ocuparán los datos centrales? Verificarás que lo más sencillo es dividir a 20 en dos, y esa posición y la siguiente ocupan los lugares centrales, ya que dejan la misma cantidad de observaciones a la izquierda y a la derecha.20/2=10. Por lo tanto, hay que buscar el dato que ocupa la posición 10 y la posición 11 en el arreglo ordenado y hacer el promedio entre ellos.

Nos conviene fijarnos en la columna de frecuencias acumuladas, aunque en este caso es sencillo darse cuenta a simple vista de que a la posición 10 la ocupa el 20 y a la 11 también la ocupa el 20. Por lo tanto, la mediana es 20.

$$\tilde{x} = 20$$

Interpretación: la mitad de los empleados tiene 20 años o menos y la otra mitad 20 años o más.

¿Notaste la diferencia con la media?, ¿a qué se debe?

Justamente porque hay un dato que se aleja en el extremo superior de la serie, que es 35. Aunque la frecuencia es poca, también está el 28, que corre la media hacia arriba. Por eso decíamos que la media es sensible a los extremos de la distribución.

Moda

Cálculo de la moda: por definición hay dos datos que se repiten más veces: el 19 y el 20.

$$\hat{x} = 19 \; y \; 20$$

Por lo tanto, esta distribución es *bimodal*.

Interpretación: Las edades de los empleados que más se repiten son 19 y 20 años.

Antes de avanzar, actualicemos las fórmulas—si corresponde—para las medidas de posición cuando los datos están en una tabla de frecuencias sin agrupar.

MEDIA | MEDIANA | MODA

Tabla 9: Resumen de las fórmulas de las medidas de posición para datos dispuestos en tablas de frecuencias sin agrupar, diferenciando población y muestra

	Fórmula	Aclaraciones
Población	$\mu = \frac{\sum(x_i \times f_i)}{\sum f_i}$	$\sum f_i = N$
Muestra	$\bar{x} = \frac{\sum(x_i \times f_i)}{\sum f_i}$	$\sum f_i = n$

Fuente: elaboración propia.

Estas fórmulas se utilizan cuando los valores que adopta la variable se repiten de tal manera que resultaría tedioso sumar la misma variable

MEDIA | **MEDIANA** | MODA

Aquí resumimos los pasos para calcular la mediana de una distribución de frecuencias para datos sin agrupar. Son los mismos pasos que se deben realizar si se trata de una población o de una muestra.

Pasos:

1. Se busca/n la/s posición/es centrales de la distribución, según si los datos son impares o pares, respectivamente.
2. En la columna de la frecuencia acumulada, se busca/n la/s posición/es central/es calculada/s en el punto 1, sabiendo que la frecuencia acumulada de una clase indica la última posición del dato correspondiente.
3. Se observa el dato que corresponde a esa posición, si el dato central es uno solo (caso n impar), o se calcula el promedio, en caso de que sean dos los datos centrales (caso n par).

MEDIA | MEDIANA | **MODA**

Para calcular la moda, no hay fórmulas en este tipo de tabla de frecuencias, solo se busca la frecuencia con mayor valor, si la hay, e inmediatamente se observa a qué variable pertenece. La variable encontrada es la moda.
Recuerda que puede haber varias modas o ninguna.

En una empresa se ha determinado el número de ausencias diarias del personal de planta y administrativo, en un lapso de 60 días hábiles. Considerando dichos valores como una muestra de 9 días tomados aleatoriamente referida a la distribución de ausencias registradas, ¿cuál es el valor de la mediana y de la moda?

0 0 1 2 0 3 3 1 0

La mediana es 1 y la moda es 0.

La mediana es 2 y la moda es 0.

La mediana es 1 y la moda es 2.

La mediana es 2 y la moda es 2.

La mediana es 0 y la moda es 1.

SUBMIT

Es el momento de volver a leer el caso de la fábrica CORLUX. Ya estamos en condiciones de calcular las medidas de posición de la primera parte del problema.

El director de Control de Calidad quería elaborar un informe con los datos de la Tabla 1 sobre las fallas en las lámparas, con base en los estadísticos de posición. Todavía no estudiamos todas las herramientas para realizar el informe, pero sí, como dijimos, las necesarias para calcular la media, la mediana y la moda de la Tabla 1.

Tengamos presente la Tabla 1 para poder realizar las operaciones convenientes.

Cálculo de la media

Recordemos la fórmula correspondiente, que se muestra en la Tabla 9:

$$\bar{x} = \frac{\sum_1^n x_i \times f_i}{\sum f_i}$$

La fórmula indica la sumatoria de los productos de cada dato por su frecuencia. Esa sumatoria se divide por el número de observaciones o, lo que es lo mismo, por la sumatoria de las frecuencias.

Una de las formas prácticas de hacerlo es agregando una columna a la tabla con el producto y sumarlo al final, es decir, vamos desglosando la fórmula.

Tabla 10: Tabla de frecuencias para las lámparas defectuosas de la empresa CORLUX

A la Tabla 1 del caso CORLUX le agregamos una columna auxiliar con el único objetivo de facilitar el cálculo para la media.

Cantidad de lámparas con algún defecto	N.° de cajas	Columna auxiliar

0	10	0 × 10 = 0
1	6	6
2	4	8
3	4	12
4	5	20
5	1	5
Totales	30	51

Fuente: elaboración propia.

Entonces la media es

$$\overline{\boldsymbol{x}} = \frac{\sum_1^n x_i \times f_i}{\sum f_i} = \frac{51}{30} = 1{,}7$$

Interpretación: Es importante interpretar que la muestra es de 30 cajas de 12 lámparas. Entonces, concluimos con que el promedio de lámparas defectuosas en la muestra es 1,7. Es relativamente bajo, por lo que tendremos que acompañar a esta medida del resto de las medidas de posición para poder dar un informe más completo.

Cálculo de la mediana

Como tenemos una tabla de frecuencias, dijimos que nos conviene agregar la columna de frecuencias acumuladas.

A la Tabla 1 del caso CORLUX le agregamos la columna de frecuencias acumuladas para poder localizar con más facilidad la mediana.

Tabla 11: Tabla de frecuencias absolutas y acumuladas para las lámparas defectuosas de la empresa CORLUX

Cantidad de lámparas con algún defecto	N.° de cajas	
0	10	10

1	6	16
2	4	20
3	4	24
4	5	29
5	1	30
Totales	30	

Fuente: elaboración propia.

Como la sumatoria de las frecuencias absolutas (n) es 30 y, por lo tanto, un número par, tenemos dos términos centrales.

Las posiciones centrales, como vimos anteriormente, se calculan mediante la fórmula (n/2) para el primer dato central y (n/2)+1 para el segundo dato. Por lo tanto, 30/2=15, y el siguiente es 16. Entonces, los datos que están en la posición 15 y 16 deben buscarse en la tabla.

La frecuencia acumulada de la segunda fila nos indica que el 1 ocupa desde la posición 11 hasta la 16 del arreglo ordenado de datos, por lo que deducimos que en la posición 15 hay un 1 y en la 16 también. Si hacemos el promedio entre ambos datos, nos da que la mediana es 1.

$$\tilde{x} = 1$$

Interpretación: La mitad de las cajas tiene una o ninguna lámpara defectuosa. La otra mitad de las cajas tiene una o más lámparas defectuosas. Esta medida debe ser acompañada con la media. Nos vamos dando cuenta de que seguramente haya valores superiores con frecuencias que elevan la media. No podemos decir, en este caso, qué medida es más certera; simplemente interpretamos cada una. Pero sí ya estamos obteniendo información y un resumen de los datos.

Cálculo de la moda

Observando la Tabla 1 o la Tabla 10, la frecuencia absoluta más alta es 10, que corresponde al valor 0. Concluimos, entonces, con que la moda es cero.

$$\hat{x} = 0$$

Interpretación de la moda: El dato que más se repite es cero, por lo que podemos aventurar que lo que se ha encontrado más veces hace referencia a lámparas sin defecto.

Para completar el informe, necesitamos estudiar las medidas de forma y si la distribución tiene algún sesgo que se pueda informar. A esto lo estudiaremos después de ver cómo se calculan las medidas de posición para datos agrupados.

Medidas de posición para datos agrupados

Cuando los datos están agrupados en intervalos, las medidas de posición se calculan con otras fórmulas que desarrollaremos a continuación, pero siempre teniendo presente lo que expresa cada una.

Determinación de la media

Para esta medida tomará especial importancia la marca de clase, el valor representativo de la clase. De otra manera, no se podría calcular, ya que a la frecuencia no la tiene una sola variable, sino un intervalo real.

Para poder determinar la media aritmética de una distribución de frecuencias, consideramos que todos los valores pertenecientes a cada intervalo están uniformemente distribuidos por dicho intervalo. De esta manera, la suma de todos ellos estará dada por el producto entre la marca de clase y la frecuencia de ese intervalo.

$$\mu = \frac{\sum(x_{mi} \times f_i)}{N}$$

Ejemplo 4

La siguiente es una distribución de frecuencias. La variable en estudio es la cantidad de dinero que una determinada empresa ha gastado en publicidad en los últimos 100 días. Tomemos los días como una población. Se desea calcular el promedio de gastos en publicidad.

Tabla 12: Gastos en dólares de una empresa en materia publicitaria

Los gastos están agrupados en intervalos y con la marca de clase en la columna siguiente. La marca de clase pertenece a la variable en estudio. Además de las frecuencias absolutas, se agregó una columna auxiliar para los cálculos que requiere la fórmula para la media.

N.° de clase	Gastos en USD [Li-Ls)	Marca de clase (xmi)	Días fi	Columna auxiliar Xmi × fi
1	250-300	275	8	2200
2	300-350	325	10	3250
3	350-400	375	15	5625

4	400-450	425	25	10 625
5	450-500	475	18	8550
6	500-550	525	10	5250
7	550-600	575	4	2300
8	600-650	625	4	2500
9	650-700	675	4	2700
10	700-750	725	2	1450
Totales	100	44 450		

Fuente: elaboración propia.

$$\mu = \frac{\sum(x_{mi} \times f_i)}{N} = \frac{44\,450}{100} = 444{,}5$$

Respuesta: En los últimos 100 días, la empresa tuvo un promedio diario de gastos en publicidad de USD 444,5.

Determinación de la mediana para datos agrupados

Al igual que la media, la mediana tiene un tratamiento distinto del que se le ha dado en los datos sin agrupar. Tiene su propia fórmula.

Aquí solo utilizaremos y explicaremos cada término de la fórmula, pero dejamos una lectura obligatoria con la justificación de su porqué.

¿Cómo calculamos la mediana cuando tenemos una tabla con intervalos?

Se realiza en dos pasos:

- **Primer paso:** calculamos la clase mediana o medial. La clase mediana es la primera clase cuya frecuencia acumulada es mayor a n/2.
- **Segundo paso:** a la clase mediana le aplicamos la siguiente fórmula:

$$Me = L_i + \frac{\Delta x\left(\frac{N}{2} - F_{(m-1)}\right)}{f_m}$$

Expliquemos cada símbolo que figura en la fórmula:

- L_i: límite inferior de la clase mediana.
- $\Delta x =$ longitud del intervalo $(ls - Li)$.
- $F_{(m-1)}$: frecuencia acumulada de la clase anterior a la clase mediana.
- f_m: frecuencia absoluta de la clase mediana.

Volvamos al ejemplo 4, tabla 12, pero ahora calculemos la mediana. Agreguemos la columna de frecuencias acumuladas.

Tabla 13: Gastos en dólares de una empresa en materia publicitaria con frecuencias absolutas y acumuladas

Los gastos están agrupados en intervalos y con la marca de clase en la columna siguiente. Además de las frecuencias absolutas, se agregó la columna de frecuencias acumuladas.

N.° de clase	Gastos en USD [Li-Ls)	Marca de clase (xmi)	Días fi	Fi
1	250-300	275	8	8
2	300-350	325	10	18
3	350-400	375	15	33
4	400-450	425	25	58
5	450-500	475	18	76
6	500-550	525	10	86
7	550-600	575	4	90
8	600-650	625	4	94
9	650-700	675	4	98
10	700-750	725	2	100
	Totales		100	

Fuente: elaboración propia.

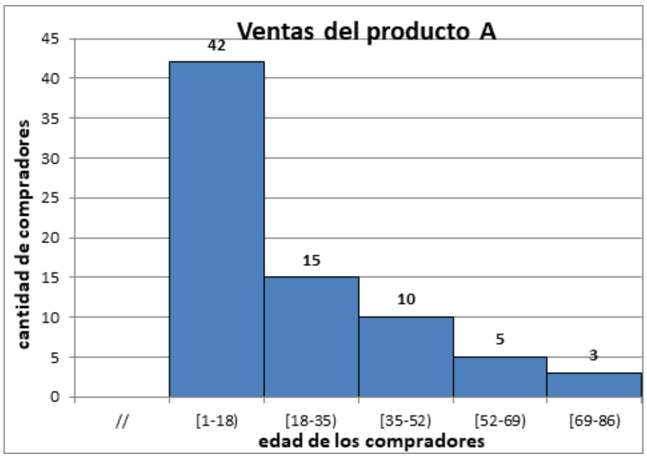
Paso 1: determinamos n/2 =100/2=50. La primera clase cuya frecuencia acumulada es mayor a 50 es la clase 4. Esta es la clase mediana, la clase que contiene a la mediana.

Paso 2: aplicamos la fórmula a la clase 4.

$$Me = L_i + \frac{\Delta x\left(\frac{N}{2} - F_{(m-1)}\right)}{f_m} = 400 + \frac{50\left(\frac{100}{2} - 33\right)}{25} = 400 + \frac{50 \times 17}{25} = 434$$

Respuesta: La mitad de los días registrados, los gastos fueron menores a USD434.La otra mitad de los días fueron mayores a USD434.

Figura 1: Histograma de ventas del producto A, de acuerdo a la edad del consumidor



En el gráfico (ver Figura 1) se muestra la cantidad de personas que compran el producto A, en una determinada ciudad, en relación con la edad del consumidor.
Se solicita determinar la clase mediana

- [1 – 18).
- [35 – 52).
- No tiene.
- [18 – 35).

SUBMIT

¿Cómo se construye la fórmula de la mediana para datos agrupados?

En este apartado, explicamos la construcción de la fórmula de la mediana para datos agrupados.

Dijimos anteriormente que la mediana está ubicada en la primera clase cuya frecuencia acumulada es mayor a n/2. Lo decimos de esta manera porque, si el número de observaciones es par, y los intervalos consecutivos, puede darse el caso, que los términos centrales queden en los extremos de dos clases distintas. Pero la clase mediana tiene que ser una sola. Por eso tomamos a la primera clase mayor a n/2, para cualquier caso, n par o impar. Si el número de observaciones es impar, se tendría un solo término central, por lo que la mediana caería sólo en una clase y no tendríamos problemas en identificarla.

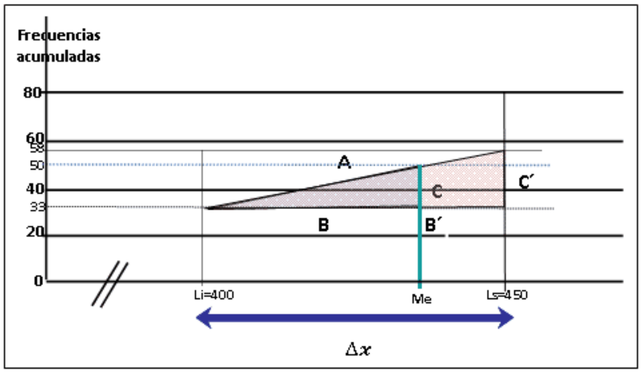
Tomemos como ejemplo la **tabla 13**, dijimos que la clase mediana es la clase 4. Significa que la mediana está entre los valores [400; 450), los datos   están en esta clase.

La frecuencia acumulada Fi de la clase mediana es 58

Analicemos con el siguiente gráfico la clase mediana.

Figura 2: gráfico de la clase mediana

En el gráfico se muestra la clase que contiene a la mediana con las frecuencias, tanto absolutas como acumuladas, con el objeto de deducir la fórmula analítica de la mediana.



En el límite inferior de la clase mediana la frecuencia acumulada es 33, Por otro lado, en el límite superior del intervalo, la frecuencia acumulada es de 58. Justamente la diferencia entre ambas frecuencias acumuladas 58-33= 25 que es la frecuencia de la clase mediana

Por otra parte, la mediana, tiene frecuencia n/2 = 50

Nos han quedado formados dos triángulos rectángulos: ABC y AB´C´ que son semejantes, por lo tanto, sus lados son proporcionales.

Entonces:

$$\frac{C'}{C} = \frac{B'}{B}$$

Pero:

$C' = 58 - 33 = 25 = f_m$ (frecuencia absoluta de la clase mediana).

$C = \frac{n}{2} - F_{m-1}$ (la posición n/2 =50 menos la Frecuencia acumulada de la clase anterior a la clase mediana).

$B' = \Delta x$ que es la amplitud del intervalo.

$B = Me - Li$ Mediana menos el límite inferior de la clase mediana.

Entonces, sustituimos

$$\frac{f_m}{\frac{n}{2} - F_{m-1}} = \frac{\Delta x}{Me - Li}$$

Y despejamos Me

$$(Me - Li).f_m = \Delta x \left(\frac{n}{2} - F_{m-1}\right)$$

$$Me = Li + \frac{\Delta x \left(\frac{n}{2} - F_{m-1}\right)}{f_m}$$

que es la fórmula citada más arriba, para calcular la mediana para datos agrupados.

Determinación de la moda para datos agrupados

También la moda tiene su propia fórmula para ser calculada. Esta fórmula surge de una interpolación en el histograma de frecuencias absolutas que dejamos más abajo como lectura obligatoria. Aquí solo utilizaremos y explicaremos cada término de la fórmula como lo hicimos con la mediana.

¿Cómo calculamos la moda cuando tenemos una tabla con intervalos?

Se realiza en dos pasos:

- **Primer paso**: calculamos la clase modal. La clase modal es la que tiene mayor frecuencia. Podemos tener una o más clases modales (o ninguna, si todas las frecuencias son iguales)
- **Segundo paso:** a cada clase modalle aplicamos la siguiente fórmula:

$$Mo = Li + \Delta x \times \frac{d_1}{(d_1 + d_2)}$$

Expliquemos cada símbolo que figura en la fórmula:

- L_i: límite inferior de la clase modal.
- Δx: longitud del intervalo (Ls-Li).
- d_1: diferencia entre la frecuencia de la clase modal y la frecuencia de la clase anterior.
- d_2: diferencia entre la frecuencia de la clase modal y la frecuencia de la clase posterior.

Observa que, esa diferencia nunca podrá ser negativa porque la frecuencia de la clase modal es la mayor. A lo sumo, puede ser cero, si hay otra clase modal consecutiva.

Retomemos el ejemplo 4, Tabla 12, pero ahora calculemos la moda.

Tabla 14: Gastos en dólares de una empresa en materia publicitaria

Los gastos están agrupados en intervalos; en este caso, solo dejamos las columnas que vamos a utilizar, que es la de frecuencias absolutas.

N.° de clase	Gastos en USD [Li-Ls)	Días fi
1	250-300	8
2	300-350	10
3	350-400	15
4	400-450	25
5	450-500	18
6	500-550	10
7	550-600	4
8	600-650	4
9	650-700	4
10	700-750	2
	Totales	100

Fuente: elaboración propia.

Paso 1: La clase modal es la 4, pues tiene la mayor frecuencia, que es 25. En este caso, coincide con la clase mediana calculada anteriormente, pero ambas no tienen por qué coincidir.

Paso 2: Aplicamos la fórmula de la moda a la clase 4.

$$Mo = Li + \Delta x \times \frac{d_1}{(d_1 + d_2)} = 400 + 50 \times \frac{25 - 15}{(25 - 15) + (25 - 18)}$$

$$Mo = 400 + 50 \times \frac{10}{10 + 7} = 400 + 50 \times \frac{10}{17} = 400 + 29{,}41 = 429{,}41$$

Respuesta: Lo que más se repite es el gasto de USD 429,41. A esto no lo podemos comprobar si no tenemos los datos recopilados. Pero es el riesgo que se corre a cambio de la practicidad de resumir los datos en una tabla con intervalos. Igualmente, lo más probable es que, si el resultado no es igual a la moda, sea uno muy cercano a ella.

Es el momento de volver a leer el caso de la fábrica CORLUX. Ya estamos en condiciones de calcular las medidas de posición de la segunda parte del problema.

El director de Control de Calidad quería elaborar un informe con los datos de la Tabla 2, sobre la vida útil de las lámparas, con base en los estadísticos de posición.

Ya vamos avanzando con las herramientas para resolver el caso, por lo que ahora tenemos que calcular la media, la mediana y la moda de la Tabla 2.

Tengamos presente la Tabla 2 para poder realizar las operaciones convenientes.

Cálculo de la media

Recordemos la fórmula correspondiente:

$$\mu = \frac{\sum(x_{mi} \times f_i)}{\sum f_i}$$

La fórmula indica la sumatoria de los productos de cada marca de clase por su frecuencia. Esa sumatoria se divide por el número de observaciones o, lo que es lo mismo, por la sumatoria de las frecuencias.

Antes tenemos que calcular las marcas de clase y agregar la columna auxiliar

$$(x_{mi} \times f_i).$$

Tabla 15: Vida útil de las lámparas muestreadas en CORLUX

La tabla muestra los registros, de menor a mayor, de las horas que duraron las lámparas muestreadas al realizar un control de calidad. Agregamos la columna de la marca de clase para calcular la media.

N.° de clase	Vida útil en horas [Li-Ls)	Marca de clase	Columna auxiliar	N.° de lámparas
1	215-515	365	365	1
2	515-815	665	1330	2
3	815-1115	965	1930	2
4	1115-1415	1265	6325	5
5	1415-1715	1565	23 475	15
6	1715-2015	1865	33 570	18
7	2015-2315	2165	47 630	22
8	2315-2615	2465	61 625	25
	Totales		176 250	90

Fuente: elaboración propia.

Entonces la media es:

$$\bar{x} = \frac{\sum(x_{mi} \times f_i)}{\sum f_i} = \frac{176\,250}{90} = 1958{,}33$$

Interpretación: La vida útil promedio de las 90 lámparas muestreadas es de 1958,33 horas por lámpara.

Cálculo de la mediana

Agregamos a la tabla la columna de frecuencias acumuladas.

Tabla 16: Vida útil de las lámparas muestreadas en CORLUX

La tabla muestra los registros, de menor a mayor, de las horas que duraron las lámparas muestreadas al realizar un control de calidad. Agregamos la columna de frecuencias acumuladas.

N.° clase	Vida útil en horas [Li-Ls)	N.° de lámparas	Fi
1	215-515	1	1
2	515-815	2	3
3	815-1115	2	5
4	1115-1415	5	10
5	1415-1715	15	25
6	1715-2015	18	43
7	2015-2315	22	65
8	2315-2615	25	90
	Totales	90	

Fuente: elaboración propia.

Paso 1: La primera clase cuya frecuencia acumulada es mayor a 90/2 =45 es la clase 7.

Paso 2:

$$Me = L_i + \frac{\Delta x\left(\frac{N}{2} - F_{(m-1)}\right)}{f_m} = 2015 + \frac{300(45-43)}{22} = 2015 + \frac{600}{22} = 2042{,}27$$

Interpretación: La mitad de las lámparas muestreadas tiene una vida útil menor a 2042,27 horas y la otra mitad tiene una vida útil mayor a 2042,27 horas.

Cálculo de la moda

Calculemos la moda observando la Tabla 16.

Paso 1: La clase modal es la 8, con 25 lámparas que tienen una vida útil entre 2315 y 2615 horas.

Paso 2:

$$Mo = Li + \Delta x \times \frac{d_1}{(d_1 + d_2)} = 2315 + 300 \times \frac{25 - 22}{(25 - 22) + (25 - 0)}$$

$$Mo = 2315 + 300 \times \frac{3}{(3 + 25)} = 2315 + 300 \times \frac{3}{28} = 2347{,}14$$

Interpretación: Se ha registrado que la duración que más se repite es de 2347,14 horas.

Otras medidas de posición

Hay otras medidas que complementan la información de la media, la mediana y la moda. Las medidas que dividen a un conjunto de datos en partes iguales son los cuartiles y los percentiles. Estos son los más utilizados, pero también están los deciles. Para poder calcularlos, los datos deben estar ordenados de menor a mayor. Ten en cuenta que los cuartiles y los percentiles son valores de la variable en estudio (como la media, la mediana y la moda).

CUARTILES

Los cuartiles son 3 valores: Q_1 , Q_2 ,que dividen a la distribución en 4 partes iguales. Q_2 coincide con la mediana.

Cada parte contiene una cuarta parte o un 25% de las observaciones.

PERCENTILES

Los percentiles son los 99 valores de la variable que dividen el arreglo ordenado de datos en 100 partes iguales.

El percentil 50 coincide con la mediana.

Lectura obligatoria: Cuartiles, deciles y percentiles

En esta lectura podrás estudiar cómo se calculan los cuartiles, los deciles y los percentiles, tanto para datos no agrupados como para datos agrupados. La lectura es muy clara y tiene ejemplos resueltos.

PDF **Cuartiles, deciles y percentiles.pdf**
518.6 KB

Fuente: Vilchis Rodríguez, Mario Arturo (2014). Cuartiles, deciles y percentiles. En Universidad Autónoma del Estado de Hidalgo. Recuperado de https://www.academia.edu/32865153/_Cuartiles_deciles_y_percentiles_

Medidas de forma. Relación entre la media, la mediana y la moda

Medidas de forma

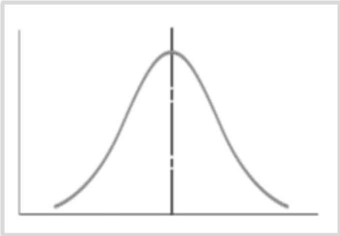
Las medidas de forma son indicadores estadísticos que nos permiten analizar la distribución de frecuencias de los datos, es decir, la forma de la distribución.

En la Lectura 2 de este módulo, habíamos estudiado que, cuando el número de intervalos tiende a infinito y la longitud del intervalo a cero, el polígono se suaviza y forma una curva. Es como si a cada dato puntual le correspondiera una frecuencia que está dada por la ordenada(altura) de la curva en ese punto.

“Las curvas que representan los datos puntuales de un conjunto de datos pueden ser simétricas o sesgadas” (Levin y Rubin, 2012, p. 58).

Figura 3: Gráfico de una curva simétrica

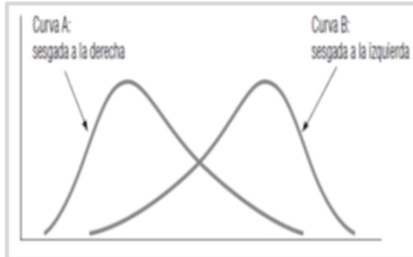
Esta curva es simétrica, ya que, al trazar una línea vertical que pase por el punto más alto, esta línea divide a la curva en dos partes iguales.



Fuente: Levin y Rubin, 2012, p. 59.

Figura 4: Comparación de dos curvas sesgadas

Estas curvas son sesgadas porque hay valores concentrados en el extremo inferior (en la curva A) y en el extremo superior (en la curva B).



Fuente: Levin y Rubin, 2012, p. 59.

Las curvas sesgadas dejan una cola hacia uno de los lados de la distribución de frecuencias.

Están sesgadas porque los valores de su distribución de frecuencias se concentran en el extremo inferior o en el superior de la escala de medición del eje horizontal. Estos valores no están igualmente distribuidos. La curva A está sesgada a la derecha (o positivamente sesgada), debido a que va disminuyendo poco a poco hacia el extremo derecho de la escala. La curva B es exactamente opuesta. Está sesgada a la izquierda (negativamente sesgada), ya que disminuye poco a poco si la recorremos hacia el extremo inferior de la escala. (Levin y Rubin, 2012, p. 59).

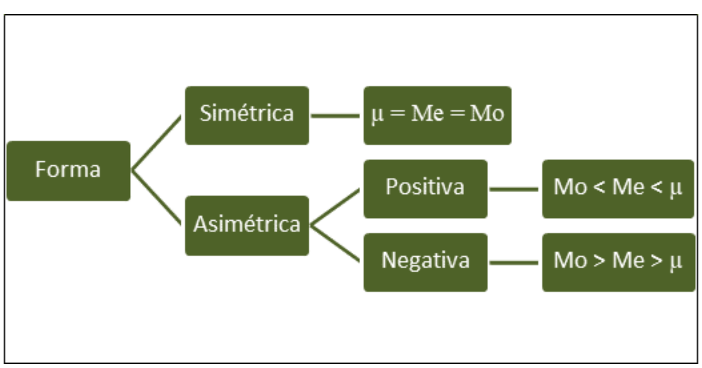
Remitimos a la página 59 de la bibliografía obligatoria, ya que se presentan casos muy interesantes en los que pueden darse estos sesgos.

Relación entre la media, la mediana y la moda

Podemos relacionar las formas de las curvas con la media, la mediana y la moda. Es más, los valores de las medidas de posición determinan la forma de la curva.

Figura 5: Relación de la media, la mediana y la moda con las formas de las curvas

En este cuadro se muestra la clasificación de las distribuciones según su forma, y su relación con las medidas de posición.



Fuente: elaboración propia.

- [ ] Si la curva es simétrica como en la Figura 1, coinciden las 3 medidas de posición.

- [ ] La curva A de la Figura 2 tiene sesgo a derecha (positiva). La moda es la menor de las 3 medidas de posición, pues está más cerca del cero en una escala tomada sobre el eje horizontal que representa a la variable en estudio. Recuerda que la moda está representada en el gráfico con el punto más alto, y su abscisa tiene un valor bajo, mientras que el sesgo a derecha indica que la media está desplazada hacia valores más altos del eje horizontal, ya que hay valores extremos con poca frecuencia, pero que la corre hacia esos valores.

- [ ] En cambio, la curva B de la Figura 2 tiene sesgo negativo. La moda adopta el mayor valor en la escala de la variable. Recuerda que la moda está representada en la gráfica con el punto más alto.

Informe final para el jefe del Departamento de Calidad de la fábrica CORLUX:

1. Con respecto a las fallas de las lámparas muestreadas, antes de salir al mercado se observa lo siguiente en la muestra de 30 cajas de 12 lámparas cada una:

   - La media aritmética es 1,7, de modo que el promedio de lámparas defectuosas en la muestra es 1,7.
   - La mediana es 1. La mitad de las cajas tiene una o ninguna lámpara defectuosa. La otra mitad de las cajas tiene una o más lámparas defectuosas.
   - La moda es 0. El dato que más se repite es cero, por lo que podemos aventurar que lo que se ha encontrado más veces hace referencia a lámparas sin defecto.

Entonces, la moda es menor que la mediana, y esta, a su vez, es menor que la media, por lo que la distribución es asimétrica con sesgo positivo. La mayor agrupación de datos está en la parte más baja de la escala. La mediana se ve un poco desplazada hacia valores más altos porque existen valores a la derecha de la distribución, pero con muy poca frecuencia.

2. Con respecto a la vida útil de las 90 lámparas muestreadas, los resultados fueron los siguientes:

   - La media aritmética es 1958,33 horas. La vida útil promedio de las 90 lámparas muestreadas es de 1958,33 horas por lámpara.
   - La mediana es 2042,27 horas. Entonces, la mitad de las lámparas muestreadas tiene una vida útil menor a 2042,27 horas y la otra mitad tiene una vida mayor a 2042,27 horas.
   - La moda es 2347,14 horas. Se ha registrado que la duración que más se repite es de 2347,14 horas.

Entonces, la moda es mayor que la mediana, y esta, a su vez, es mayor que la media, por lo que la distribución es asimétrica con sesgo negativo. La mayor agrupación de datos está en la parte más alta de la escala. La media se ve un poco desplazada hacia valores más bajos porque existen valores a la izquierda de la distribución, pero con muy poca frecuencia.

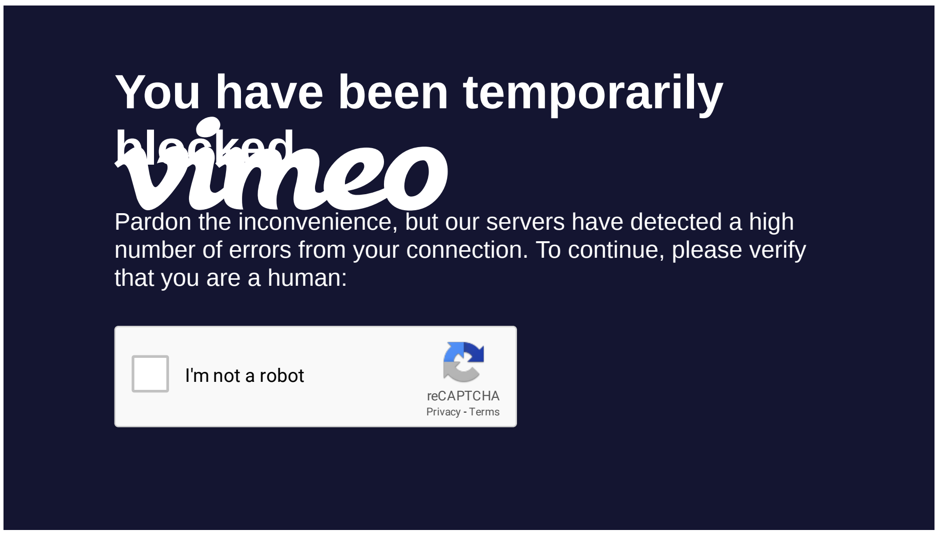
Observa que el informe es objetivo, la toma de decisiones está a cargo del jefe del Departamento de Calidad. En la próxima lectura, veremos que pueden completarse más estos informes con las medidas de dispersión, tan importantes en los procesos productivos.

Video conceptual

Video conceptual 1: Parámetros de tendencia central. Media. Mediana y Moda

En el video además de hacer un breve repaso sobre las medidas de tendencia central, se comparan las tres medidas y se estudian: la simetría o las asimetrías de una distribución y las condiciones que deben darse para que estas ocurran.

VIMEO



VC - Herramientas Matemáticas III - Estadística I - Módulo 1 (CANVAS)

VER EN VIMEO >

Referencias

Levin, R. y Rubin, D. (2012). *Estadística para Administración y Economía* (7.a ed.). Naucalpan de Juárez, MX: Pearson.

Berenson, M., Levine, D. y Krehbiel, T. (2006). *Medidas numéricas descriptivas. En Autores, Estadística descriptiva* (pp. 72-121). Naucalpan de Juárez, MX. Pearson.

Anderson, D. R., Sweeney, D. J. y Williams, T. A. (2008). *Estadística Descriptiva: Medidas numéricas. En Autores, Estadística para Administración y Economía* (pp. 89-109). D.F., MX: Cengage Learning.