Universidad del Cauca

ISO 9001

ISO 9001:2015 SC- CER 450832

IQNet: CO- SC-CER450832

Una Acreditación con Rostro Humano

Hacia una Universidad comprometida con la paz territorial

Estadística

Ingeniería de Sistemas

Indicadores estadísiticosClase #4

Son números que caracterizan el comportamiento del conjunto de datos. Dos de estos números son de particular importancia para los responsables de tomar decisiones: la **tendencia central** y la **dispersión**.

Indicadores de tendencia central

También llamadas medidas de ubicación, o promedios, son números que definen cual es el valor al rededor del cual se concentran los datos de una población o una muestra.
Un promedio es un valor típico o representativo de un conjunto de datos, motivo por el cual tiende a concentrarse en el centro del conjunto de datos ordenados de acuerdo con su magnitud.

Media Poblacional (μ): sean $x_1, x_2, \ldots, x_N$ los N valores observados en la población.

$$\mu = \sum_{i=1}^{N} \frac{x_i}{N} = \frac{x_1 + x_2 + \cdots + x_N}{N}.$$

Media Muestral ($\bar{x}$): Sea $x_1, x_2, \ldots, x_n$ una muestra extraída de determinada población.

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

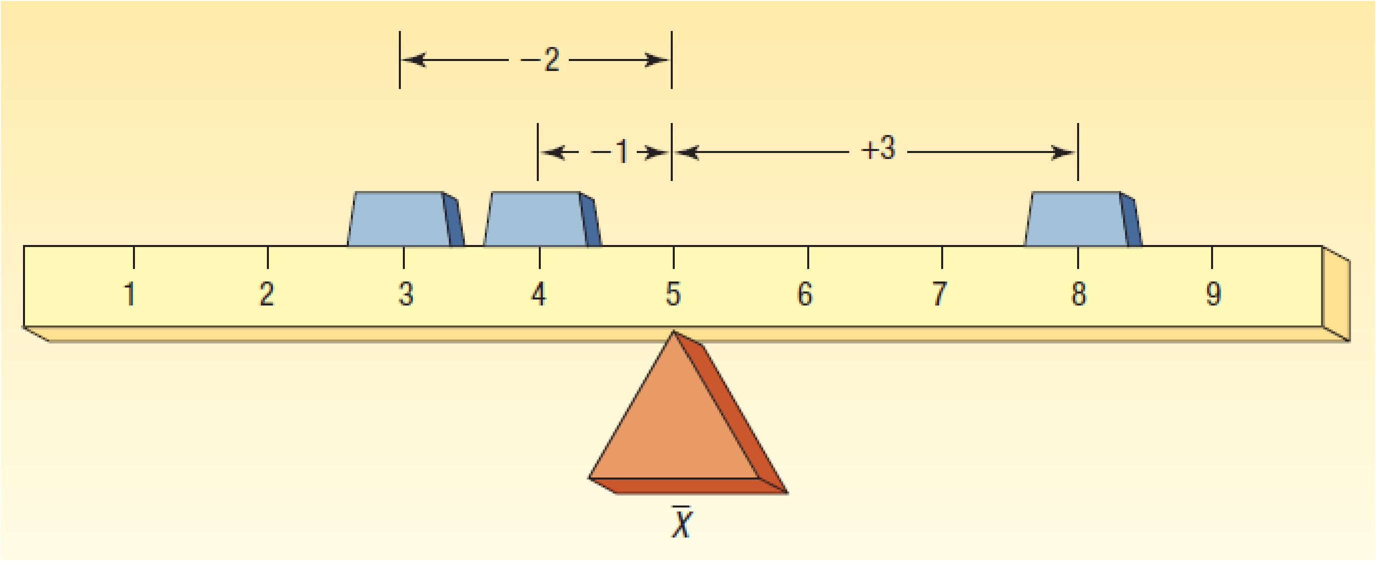
Observación.

1. Cualquier característica medible de una población recibe el nombre de **Parámetro**, mientras que cualquier característica medible de una muestra se denomina **Estadístico**.
2. Las dos expresiones matemáticas que definen a la media poblacional y la media muestral, recibe el nombre de **Media Aritmética**.

Propiedades.

1. Todo conjunto de datos posee una media.
2. Todos los valores de la población (o muestra) se incluyen en el cálculo de la media.
3. La media es única.
4. La suma de las desviaciones (diferencia entre la media y cada valor) de cada valor a la media es cero, esto es:

$$\sum_{i=0}^{n}(x_i - \bar{x}) = 0.$$



Ejemplo.

Considere los datos 2, 6, 11, 8, 11, 4, 7, 5. los cuales constituyen una muestra de tamaño 8.

$$\bar{x} = \frac{2+6+11+8+11+4+7+5}{8} = 6.75$$

Observación.

La media tiene un punto débil. Recuerde que, por la Propiedad 2. el valor de cada elemento de una muestra, o población, se utiliza cuando se calcula la media. Si uno o dos de estos valores son extremadamente grandes o pequeños comparados con la mayoría de los datos, la media podría no ser un promedio adecuado para representar los datos.

Suponga que tenemos un dato más en los datos del ejemplo anterior dado por 90. En este caso tendríamos que:

$$\bar{x} = \frac{2+6+11+8+11+4+7+5+90}{9} = 16.$$

Media Ponderada ($\bar{x}_w$)**:** Considere una muestra de tamaño $w = w_1 + w_2 + \cdots + w_n$: $x_1, x_2, \ldots, x_w$ y suponga que:

x_1 se repite w_1 veces,

x_2 se repite w_2 veces,

$\vdots$

x_n se repite w_n veces.

entonces la media ponderada de ese conjunto de datos se define como:

$$\bar{x}_w = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i} = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n}.$$

Ejemplo.

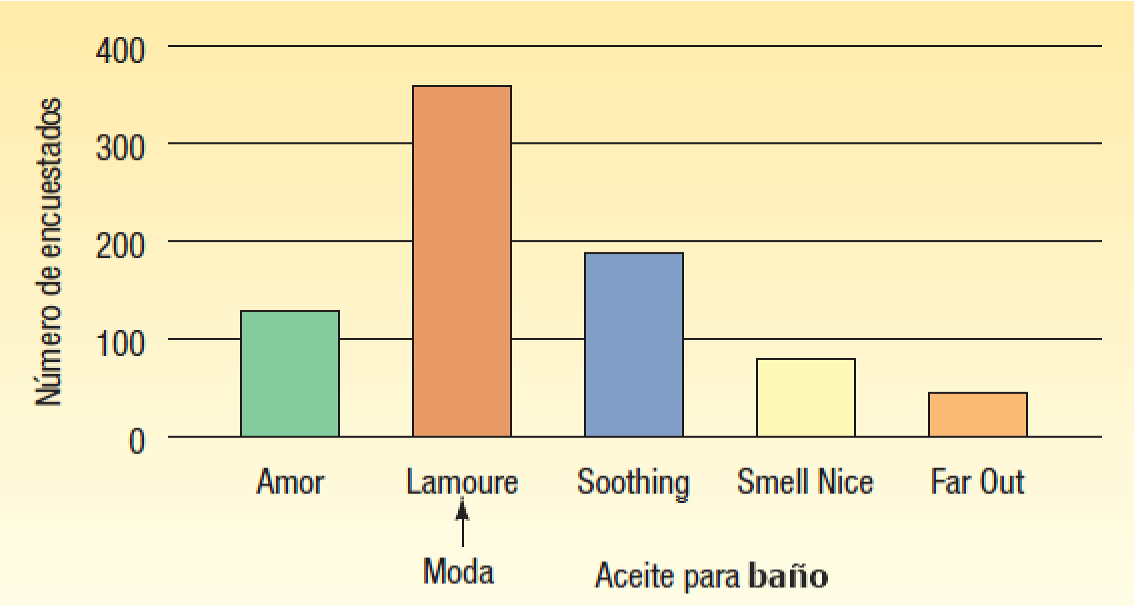
Suponga que un restaurante vende refrescos medianos, grandes y gigantes a \$900, \$1250 y \$1500. De las 10 últimas bebidas que se vendieron 3 eran medianas, 4 grandes y 3 gigantes. ¿Cuál es el precio promedio de venta de las últimas 10 bebidas?

Moda (M_o): es el dato que ocurre con mayor frecuencia en una muestra o población. En general puede ser que no exista la moda y también es posible que exista más de una moda.

La moda del conjunto de datos $\{2, 2, 5, 7, 9, 9, 9, 10, 10\}$ es $M_o = 9$.

El conjunto de datos $\{3, 5, 8, 10, 12, 15\}$ no tiene moda.

El conjunto de datos $\{2, 3, 4, 4, 4, 5, 5, 7, 7, 7, 9\}$ tiene dos modas $M_1 = 4$ y $M_2 = 7$.



Ejemplo.

Una compañía creó cinco aceites para baño. La siguiente gráfica muestra los resultados de una encuesta de mercado que se diseñó para determinar qué aceite para baño prefieren los consumidores.

Mediana Muestral ($\tilde{x}$): Es el valor ubicado en el centro de los datos ordenados. Considere una muestra de tamaño n: $x_1, x_2, \ldots, x_n$, y sean $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ los elementos de la muestra ordenada:

$$\tilde{x} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{si } n \text{ es impar} \\ \frac{1}{2}\left[x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right], & \text{si } n \text{ es par} \end{cases}$$

Observación.

Si los datos contienen uno o dos valores muy grandes o muy pequeños, la media aritmética no resulta ser un promedio representativo de los datos. En tal caso es posible y conveniente describir el centro de dichos datos a partir de la mediana.

Ejemplo.

Suponga que busca un condominio en una zona exclusiva de Bogotá. Su agente de bienes raíces le dice que el precio típico de las unidades disponibles en este momento es de $110 000 (dólares).

¿Aún insiste en seguir buscando?

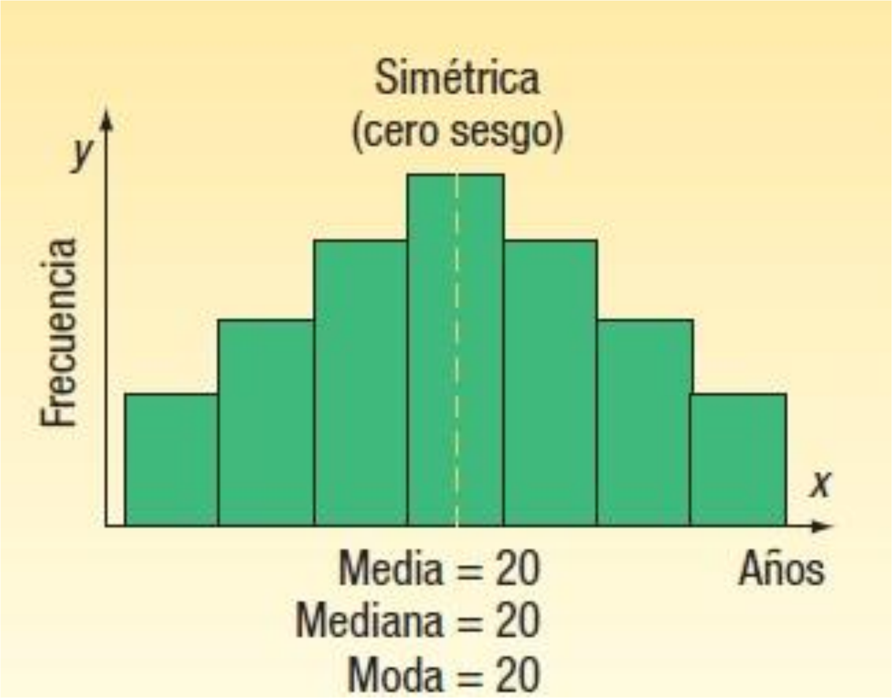
Si usted se ha fijado un presupuesto máximo de $75.000, podría pensar que los condominios se encuentran fuera de su presupuesto.

Los costos son de $275.000 en el caso de un lujoso penthouse, y de $70.000, $60.000, $80.000, $65.000. **El promedio aritmético de estos valores es de $110 000.**

Parece que un precio de poco más o menos $70,000 es un promedio más típico o representativo, y así es.

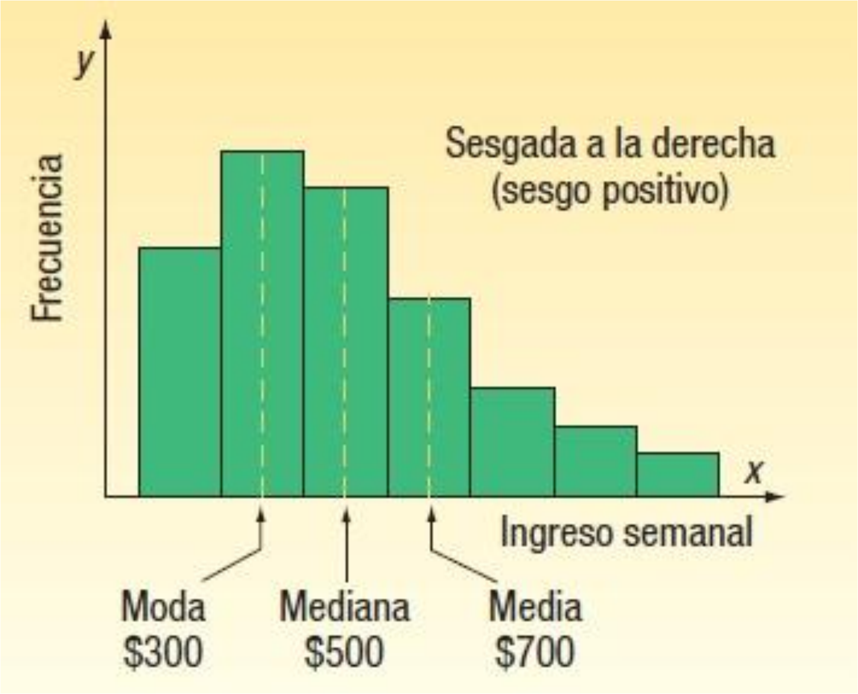
En casos como éste, la mediana proporciona una medida de ubicación más adecuada.

Comparación de la media, la mediana y la moda



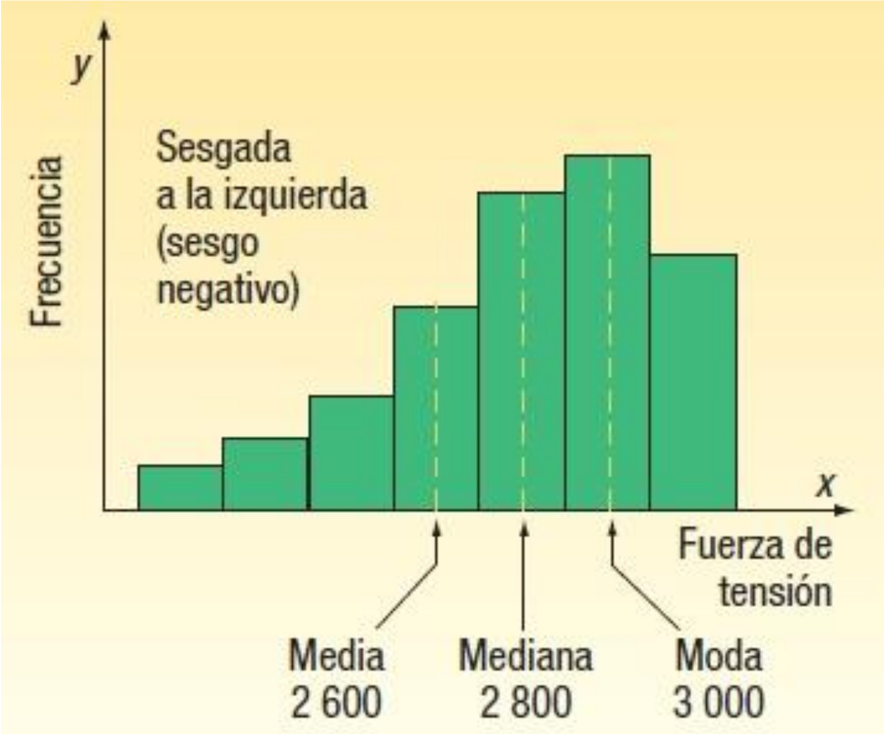
Las distribuciones simétricas que sólo contienen una moda siempre tienen el mismo valor para la media, la mediana y la moda, es decir:

$$\bar{x} = M_o = \tilde{x}$$



En una distribución con sesgo positivo, la moda todavía se encuentra en el punto más alto de la distribución, la mediana está a
la derecha de la moda y la media se encuentra todavía más a la derecha de la moda y la mediana.

$$M_o \leq \tilde{x} \leq \bar{x}$$



En una distribución con sesgo negativo, la moda sigue siendo el punto más alto de la distribución, la mediana está a la izquierda y la media se encuentra todavía más a la izquierda de la moda y la mediana.

$$\bar{x} \leq \tilde{x} \leq M_o$$

Cuando la población está sesgada negativa o positivamente, la mediana suele ser la mejor medida de posición, debido a que siempre está entre la moda y la media. La frecuencia de ocurrencia de un solo valor no influye mucho en la mediana como es el caso de la moda, ni la distorsionan los valores extremos como la media.

¡Gracias por su atención!

Universidad del Cauca

www.unicauca.edu.co

Ejercicios

1. Investigue sobre la media geométrica y armónica, y sobre la relación que existe entre estas y la media arimética. Verifique dicha relación con la muestra de datos: $\{7, 5, 7, 3, 7, 4\}$.

2. Calcule la media de las siguientes muestras de valores, y verifique que se cumple $\sum (x - \bar{x}) = 0$ Muestra 1: $\{5, 9, 4, 10\}$. Muestra 2: $\{1.3, 7.0, 3.6, 4.1, 5.0\}$.

3. Suponga que va a la tienda y gasta \$61.85 en 14 artículos. ¿Cuál es el precio promedio por artículo?

4. En junio, una inversionista compró 300 acciones de Oracle (una compañía de tecnología de la información) a \$20 cada una. En agosto compró 400 acciones más a \$25. En noviembre compró otras 400 acciones, pero el precio bajó a \$23 cada título. ¿Cuál es el precio promedio ponderado de cada acción?

5. Una librería especializada que se dedica a la venta de libros usados por internet. Los libros de pasta blanda cuestan \$1.00 (en dolares) cada uno y los de pasta dura, \$3.50 cada uno. De los 50 libros que se vendieron el pasado martes por la mañana, 40 eran de pasta blanda y el resto de pasta dura. ¿Cuál fue el precio promedio ponderado de un libro?

En los ejercicios 6. 7. y 8. determine la media aritmética, la moda y la mediana. Con estos resultados determine si la distribución de los datos es simétrica, casi simétrica, o sesgada, luego con un histograma de frecuencias verifique sus conclusiones.

6. Los siguientes son los números de cambios de aceite de los últimos 7 días en un taller automotriz de la ciudad de Popayán.

41	15	39	54	31	15	33

7. El siguiente es el cambio porcentual en el ingreso neto del año pasado al presente en una muestra de 12 compañías constructoras de Colombia.

5	1	-10	-6	5	12	7	8	2	5	-1	11

8. Las siguientes son las edades de 10 personas que se encuentran en la sala de videojuegos de un centro comercial de la ciudad de Cali a las 10 de la mañana.

12	8	17	6	11	14	8	17	10	8

9. La tasa de desempleo en Colombia durante los 12 meses de 2004 aparece en la siguiente tabla:

Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
8. 7	8.8	8.7	7.8	7.3	7.8	6.6	6.5	6.5	6.8	7.3	7.6

(a) ¿Cuál es la media aritmética de la tasa de desempleo en Colombia?

(b) Encuentre la mediana y la moda de la tasa de desempleo.

(b) Calcule la media aritmética y la mediana sólo de los meses de invierno (de diciembre a marzo). ¿Es muy diferente con respecto a los resultados obtenidos de la población total?

10. Para medir el índice de la calidad del aire en la ciudad de Bogotá se utiliza la siguiente escala: Buena (0–50), Regular (51–100), Mala (101–150), Muy mala (151–200), extremadamente mala (201–300), y Peligrosa (301–500).

    En la Tabla 1.15 se muestran los niveles de contaminación promedio mensual por ozono (O_3), medido en partes por millón (ppm), de los 10 primeros meses del año 2017.

Tabla 1.15. Promedios mensuales de contaminación por ozono (O_3).

Mes	Nivel de O_3
Enero	87
Febrero	93
Marzo	85
Abril	95
Mayo	128
Junio	90
Julio	86
Agosto	83
Septiembre	64

(a) Calcula la media, la mediana y la moda de este conjunto de datos e interpreta los resultados.

(b) ¿De qué manera el valor del nivel de contaminación del mes de mayo influye en los valores de la media y la mediana?