open green road

Guía Matemática

Medidas de Tendencia Central

tutora: Jacky Moreno

puntaje nacional.cl

1. Medidas de Tendencia Central

En Chile existen numerosas instituciones que realizan investigaciones estadísticas sobre diferentes temas, tales como la salud pública, la educación, la cultura, la locomoción colectiva, entre otros.

En tu caso cuando en el colegio te entregan una nota 5,0 en una prueba, tiendes a preguntar cómo les fue a tus amigos para así saber que tan variadas fueron las notas del curso, esto te permite saber que tan buena o mala fue tu nota 5,0 ya que la nota por si sola tiende a carecer de significado. En situaciones como éstas, de forma intuitiva, tiendes a realizar un estudio sobre las notas de tu curso de forma tal que puedas decir frases como "me fue súper bien en la prueba, estuve por sobre el promedio de mi curso", "la prueba estuvo difícil, fui la nota más baja" o "soy del montón, la mayoría de mis compañeros obtuvo nota 5,0".

De acuerdo a lo planteado anteriormente, es que nos damos cuenta que no basta con recolectar, organizar y presentar la información en una tabla estadística o en un gráfico, sino que debemos contar con otros elementos de referencia que nos permitan analizar la información desde otra perspectiva. En este caso acudiremos a las medidas de tendencia central para interpretar el comportamiento de los datos.

*Las **medidas de tendencia central** son valores numéricos que expresan el grado de centralización de los datos que representan.*

Las medidas de tendencia central más utilizadas son:

1.1. Media Aritmética ($\overline{X}$)

La media aritmética, también denominada promedio, es una medida de tendencia central que sólo se puede aplicar en variables cuantitativas.

*La **media** se define como la suma de los valores de todas las observaciones divididos por el número total de datos.*

Algunas ideas sobre esta medida de tendencia central son:

- No es necesario que los datos estén ordenados para calcular la media aritmética.
- Todos los datos son incluidos en el cálculo de la media aritmética.
- Un conjunto de datos solo tiene una media aritmética.
- El valor numérico puede o no coincidir con algunos de los datos del conjunto.
- Se utiliza generalmente para comparar dos o más conjuntos de datos.
- Es sensible a una distribución muy asimétrica de los datos, es decir, pierde precisión cuando hay valores extremos, muy grandes o muy pequeños, en comparación con el general de la muestra.
- Cuando se aplica en datos agrupados en intervalos, la medida pierde precisión debido a que existe una pérdida de información al agrupar los datos en clases.

A continuación mostraremos como calcular la media aritmética en distintas situaciones de acuerdo a como se nos presentan los datos:

1. Con miras a las compras previas a las fiestas de fin de año, el Servicio Nacional del Consumidor (SERNAC) decide realizar un sondeo para conocer el costo de una cena familiar. La muestra se tomó entre el 11 y 13 de Diciembre del 2012 y los resultados que arrojó en cuanto al precio del producto "Duraznos mitades, grado 2, Dos Caballos" en diferentes sectores de Santiago se muestran a continuación:

$$\$849 \ - \ \$856 \ - \ \$889 \ - \ \$854 \ - \ \$907$$

En esta situación se nos presentan los datos por extensión, por lo tanto, basta con sumar cada dato y dividirlo por el total de éstos para obtener la media aritmética.

$$\overline{X} = \frac{849 + 856 + 889 + 854 + 907}{5}$$
$$\overline{X} = 871$$

En base al desarrollo anterior, podemos decir que el valor promedio del producto "Duraznos mitades, grado 2, Dos Caballos" es de \$871.

De forma general, cuando se nos presentan los **datos por extensión**, el método para encontrar la media aritmética consiste en:

a) Sumar todos los datos (x_i).

b) Dividir el resultado de la suma en el total de datos (n).

De esta forma la media aritmética es:

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots + x_k}{n}$$

2. Una organización que promueve la vida saludable decide estudiar los kilogramos de fruta que compran 120 familias. Durante una semana se registraron los kilogramos de fruta que compraron cada familia, obteniéndose los siguientes datos tabulados:

Kilos	**Familias**
1	5
2	18
3	26
4	17
5	12
6	34
7	8

En esta situación nos entregan los datos tabulados con su respectiva frecuencia, por lo tanto en vez de sumar 5 veces el número uno, acudiremos a la multiplicación, de esta forma en vez de operar $1 + 1 + 1 + 1 + 1$ reduciremos la expresión a $1 \cdot 5$. Al realizar lo mismo con todos los datos de la tabla obtenemos la media aritmética de la siguiente forma:

$$\overline{X} = \frac{1 \cdot 5 + 2 \cdot 18 + 3 \cdot 26 + 4 \cdot 17 + 5 \cdot 12 + 6 \cdot 34 + 7 \cdot 8}{5 + 18 + 26 + 17 + 12 + 34 + 8}$$
$$\overline{X} = \frac{507}{120}$$
$$\overline{X} = 4,225$$

A partir del resultado anterior podemos decir que las familias consumen $4,225$ kilogramos de fruta en promedio a la semana.

De forma general, en una **tabla de distribución de frecuencia**, el método para encontrar la media aritmética consiste en:

a) Multiplicar cada dato (x_i) por su frecuencia(f_i).

b) Sumar todos los resultados anteriores.

c) Dividir el resultado de la suma en el total de datos (n).

De esta forma la media aritmética es:

$$\overline{X} = \frac{x_1 \cdot f_1 + x_2 \cdot f_2 + \ldots + x_k \cdot f_k}{n}$$

3. Para que un estudiante pueda entrar a la Universidad de Chile a estudiar Ingeniería y Ciencias-Plan Común, debe postular con un puntaje PSU que se obtiene de acuerdo al valor que tiene cada puntaje para la carrera. A continuación se muestran los puntajes obtenidos por un estudiante en el proceso de admisión 2013:

Requisitos	**Puntaje**	**Ponderación** [%]
Notas enseñanza media	723	10
Ranking	770	10
Prueba Lenguaje	655	10
Prueba Matemática	746	50
Prueba Ciencias	708	20

En este caso para obtener el puntaje con el cuál el estudiante postula a la carrera universitaria se debe calcular la media aritmética ponderara ya que no todos los datos tienen igual valor.

$$\overline{X} = \frac{723 \cdot 10 + 770 \cdot 10 + 655 \cdot 10 + 746 \cdot 50 + 708 \cdot 20}{10 + 10 + 10 + 50 + 20}$$

$$\overline{X} = \frac{7.230 + 7.700 + 6.550 + 37.300 + 14.160}{100}$$
$$\overline{X} = \frac{72.940}{100}$$
$$\overline{X} = 729,4$$

En base a lo anterior, el estudiante postula con 729,4 puntos al Plan Común de Ingeniería y Ciencias de la Universidad de Chile.

De forma general, en una **tabla con datos ponderados**, el método para encontrar la media aritmética consiste en:

a) Multiplicar la ponderación (p_i) por su dato(x_i).
b) Sumar todos los resultados anteriores.
c) Dividir el resultado de la suma total de las ponderaciones (P).

De esta forma la media aritmética es:

$$\overline{X} = \frac{x_1 \cdot p_1 + x_2 \cdot p_2 + \ldots + x_k \cdot p_k}{P}$$

4. En un hospital público se desea estudiar el tiempo de espera de los pacientes entre las 22:00 horas y las 00:00 horas. Durante un día viernes se registraron los tiempos de espera de los pacientes, obteniéndose los siguientes datos tabulados:

Tiempo [min]	Pacientes	Marca de clase
[0-30[	5	15
[30-60[	8	45
[60-90[	6	75
[90-120[	15	105
[120-150[	35	135
[150-180[	12	165
[180-210[	20	195
[210-240[	24	225
[240-270[	50	255
[270-300[	40	285

En esta situación no conocemos los datos recolectados por el hospital sino que solo conocemos los intervalos en los que estos están agrupados, por lo tanto, para calcular la media aritmética haremos uso de la marca de clase[1] debido a que éste es un valor representativo de cada intervalo.

$$\overline{X} = \frac{5 \cdot 15 + 8 \cdot 45 + 6 \cdot 75 + 15 \cdot 105 + 35 \cdot 135 + 12 \cdot 165 + 20 \cdot 195 + 24 \cdot 225 + 50 \cdot 255 + 40 \cdot 285}{5 + 8 + 6 + 15 + 35 + 12 + 20 + 24 + 50 + 40}$$

$$\overline{X} = \frac{75 + 360 + 450 + 1.575 + 4.725 + 1.980 + 3.900 + 5.400 + 12.750 + 11.400}{215}$$

$$\overline{X} = \frac{42.615}{215}$$

$$\overline{X} \approx 198,21 \tag{1}$$

[1] Recordar que la marca de clase corresponde al promedio entre los límites de la clase o intervalo.

De acuerdo al cálculo anterior, el tiempo de espera promedio por los pacientes del hospital público fue de 198,21 minutos.

De forma general, en una **tabla de frecuencia con datos agrupados en intervalos**, el método para encontrar la media aritmética consiste en:

a) Multiplicar la marca de clase (m_i) de cada intervalo por su frecuencia(f_i).

b) Sumar todos los resultados anteriores.

c) Dividir el resultado de la suma en el total de datos (n).

De esta forma la media aritmética es:

$$\overline{X} = \frac{m_1 \cdot f_1 + m_2 \cdot f_2 + \cdots + m_k \cdot f_k}{n}$$

✍ *Ejercicios* 1

1. Un grupo de 8 amigos deciden ir a un pub ubicado en Bellavista. Si la cantidad promedio de cervezas que se tomaron por persona fue de 4 latas, siendo que 3 de ellos tomaron 2, 5 y 6 latas respectivamente, ¿cuál es la cantidad de cerveza promedio que se tomaron los otros 5 amigos?

2. El promedio de horas de estudio de un grupo de 12 estudiantes es de 10 horas a la semana. Si se incorpora al grupo un estudiante que dedica 5 horas para estudiar a la semana, ¿cuál es el promedio de horas de estudio del nuevo grupo de estudiantes?

3. Afuera del Hospital Sotero del Rio 3 personas suben a un colectivo haciendo que la masa promedio de las personas que están dentro de éste sea de 325[kg]. Si 20 minutos después se baja la primera persona del colectivo entonces la masa promedio se reduce a 210[kg]. ¿Cuál es la masa de la persona que se bajó primero del colectivo?

4. Si la estatura promedio de un grupo de 15 jóvenes es de $1,63$[m] y la de otro grupo de 25 jóvenes es de $1,60$[m]. ¿Cuál es la estatura promedio entre los dos grupos?

5. La media aritmética de Julieta en el examen de inglés fue de 638 puntos. El examen evaluaba 3 áreas: escritura, lectura y audición, siendo el valor de cada área de un 35 %, 25 % y 40 % respectivamente. Si Julieta obtuvo 510 puntos y 715 puntos en las dos primeras áreas, ¿cuántos puntos obtuvo en el área de audición?

1.2. Mediana (M_e)

La mediana es una medida de tendencia central que es aplicada sólo en variables cuantitativas.

*La **mediana** se define como el valor numérico que divide a un conjunto de datos, ordenados de manera creciente o decreciente, en dos partes iguales, es decir, deja por debajo y por encima de sí el 50 % de la distribución de datos.*

Algunas ideas sobre esta medida de tendencia central son:

- Es necesario que los datos estén ordenados para calcular la mediana.
- Un conjunto de datos sólo tiene una mediana.
- El valor numérico puede o no coincidir con algunos de los datos del conjunto.
- Es estable a los valores extremos de un conjunto de datos.

A continuación mostraremos como calcular la mediana en distintas situaciones de acuerdo a como se nos presentan los datos:

1. El Servicio Nacional del Consumidor (SERNAC) decide realizar un sondeo para conocer el precio del combustible líquido según sector de la Región Metropolitana. El estudio se llevo a cabo el día 31 de Diciembre del 2012 y los resultados registrados sobre el precio promedio de la gasolina de 97 octanos se presentan a continuación:

   $$\$780 \ - \ \$774 \ - \ \$792 \ - \ \$771 \ - \ \$776$$

   En este caso el número de datos es impar, por lo tanto, para calcular la mediana basta con ordenar los datos de forma creciente o decreciente y determinar el valor central.

   $$\$771 \ - \ \$774 \ - \ \boxed{\$776} \ - \ \$780 \ - \ \$792$$

   De acuerdo al desarrollo anterior, la mediana del precio de la gasolina de 97 octanos es \$776, valor que corresponde al tercer lugar (X_3) de la distribución ordenada de datos.

   De forma general, cuando se nos presenta un **conjunto impar de datos discretos por extensión**, el método para encontrar la mediana consiste en:

   a) Ordenar los datos de forma creciente o decreciente.
   b) Localizar el valor que divide en dos partes iguales al total de datos (n).

   De esta forma la mediana es el dato que ocupa el lugar:

   $$M_e = X_{\frac{(n+1)}{2}}$$

2. El Servicio Nacional del Consumidor (SERNAC) realizó un sondeo sobre el precio del pan durante el año 2012 en la Región Metropolitana. Los registros tomados en el mes de Septiembre para 6 tipos de panes se muestran a continuación:

   $$\$962 \ - \ \$912 \ - \ \$1.239 \ - \ \$1.174 \ - \ \$1.342 \ - \ \$1.325$$

   En esta situación el número de datos es par, por lo tanto, para calcular la mediana tendremos que calcular el promedio entre los dos datos centrales que tengamos luego de ordenar nuestra información de forma creciente o decreciente.

   $$\$912 \ - \ \$962 \ - \ \boxed{\$1.174} \ - \ \boxed{\$1.239} \ - \ \$1.325 \ - \ \$1.342$$

$$M_e = \frac{1.174 + 1.239}{2}$$
$$M_e = \frac{2.413}{2}$$
$$M_e = 1.206,5$$

De acuerdo al desarrollo anterior, la mediana del precio del pan en la Región Metropolitana es de $\$1.206,5$.

De forma general, cuando se nos presenta un **conjunto par de datos discretos por extensión**, el método para encontrar la mediana consiste en:

a) Ordenar los datos de forma creciente o decreciente.

b) Localizar los dos valores centrales ($X_{n/2}$ y $X_{(n/2)+1}$) de la distribución total de datos (n).

c) Calcular el promedio entre los dos valores encontrados anteriormente.

De esta forma la mediana es:

$$M_e = \frac{X_{n/2}+X_{(n/2)+1}}{2}$$

3. El 15 de Diciembre se realizó la corrida Nike "We run Santiago 10K 2012". El tiempo que se demoraron en recorrer los primeros 5 kilometros 72 mujeres entre 16 años y 19 años se encuentra registrado en la siguiente tabla:

Tiempo [min]	**Mujeres**
[20-25[	1
[25-30[	8
[30-35[	23
[35-40[	30
[40-45[	6
[45-50[	4

En una **tabla de frecuencia con datos agrupados en intervalos**, el método para encontrar la mediana consiste en:

a) Determinar el valor numérico de la mitad de los datos ($n/2$)

b) Localizar el intervalo en el cuál esta contenido ese valor.

c) Sustituir los siguientes valores:

- n =Número de observaciones.
- a =Amplitud del intervalo seleccionado.
- L_i =Límite inferior del intervalo seleccionado.
- f_i =Frecuencia absoluta del intervalo seleccionado.

- F_i =Frecuencia absoluta acumulada del intervalo anterior a la clase que contiene a la mediana.

En la expresión:

$$M_e = L_i + a \cdot \left[\frac{\left(\frac{n}{2}\right) - F_i}{f_i} \right]$$

En esta situación tenemos un total de 72 mujeres. Para saber en qué intervalo esta nuestra mediana dividimos el total de datos en 2:

$$\frac{72}{2} = 36$$

De acuerdo a lo anterior, nuestra mediana corresponde al dato número 36, que está ubicado de la clase $[35 - 40[$.

Luego de tener nuestro intervalo identificado, determinamos los valores numéricos de los datos necesarios para obtener la mediana:

- $n = 72$
- $a = 40 - 35 = 5$
- $L_i = 35$
- $f_i = 30$
- $F_i = 1 + 8 + 23 = 32$

Finalmente sustituimos en la expresión:

$$M_e = L_i + a \cdot \left[\frac{\left(\frac{n}{2}\right) - F_i}{f_i} \right]$$

$$M_e = 35 + 5 \cdot \left[\frac{\left(\frac{72}{2}\right) - 32}{30} \right]$$

$$M_e = 35 + 5 \cdot \left[\frac{4}{30} \right]$$

$$M_e = 35 + \frac{2}{3}$$

$$M_e = \frac{107}{3}$$

$$M_e \approx 35,67$$

La mediana de las mujeres que corrieron el evento “We run Santiago 10K 2012” es de 35,67 minutos.

1.3. Moda (M_o)

La moda es una medida de tendencia central que es aplicada en variables cuantitativas y variables cualitativas.

*La **moda** se define como el dato que posee mayor frecuencia absoluta, es decir, el valor que más se repite.*

Algunas ideas sobre esta medida de tendencia central son:

- No es necesario que los datos estén ordenados para calcular la moda.
- Un conjunto de datos puede tener más de una moda o puede que este presente.
- El valor numérico coincide con algún dato del conjunto.

A continuación mostraremos como calcular la moda en distintas situaciones de acuerdo a como se nos presentan los datos:

1. El Servicio Nacional del Consumidor (SERNAC) decide realizar el sondeo mensual de precios de los medicamentos en farmacias del área metropolitana correspondiente a Noviembre del 2012. Los precios del medicamento "Clorfenamina Maleato, 4 mg de 20 Grageas, Laboratorio Chile" registrados en distintas farmacias se muestran a continuación:

   $$\$190 \; - \; \$195 \; - \; \$290 \; - \; \$290 \; - \; \$350 \; - \; \$240$$

   En este caso para encontrar la moda debemos determinar el valor que más se repite dentro de los datos.

   $$\$190 \; - \; \$195 \; - \; \$290 \; - \; \$290 \; - \; \$350 \; - \; \$240$$

   En este caso la moda es \$290 con una frecuencia de 2.

2. El 15 de Diciembre se realizó la corrida Nike "We run Santiago 10K 2012". El tiempo que se demoraron en realizar la corrida los 30 participantes hombres entre 60 años y 64 años se encuentran registrados en la siguiente tabla:

Tiempo [min]	**Hombres**
[40-50[	2
[50-60[	10
[60-70[	13
[70-80[	4
[80-90[	1

   Al presentarnos la información a través de una tabla de frecuencia con datos agrupados en intervalos, se acepta como válido que la moda corresponda a la marca de clase del intervalo que posea mayor frecuencia absoluta.

En esta situación, el intervalo que tiene mayor frecuencia absoluta corresponde a $[60, 70[$ con 13 adultos que realizaron la carrera dentro de ese tiempo. De acuerdo a lo descrito arriba la moda sería:

$$M_o = \frac{60+70}{2}$$
$$M_o = \frac{130}{2}$$
$$M_o = 65$$

En este caso la moda es de 65 minutos. Sin embargo, cuando se desea mayor precisión se puede trabajar de la siguiente forma:

a) Localizar el intervalo que posee mayor frecuencia absoluta.

b) Sustituir los siguientes valores:

- a =Amplitud del intervalo seleccionado.
- L_i =Límite inferior del intervalo seleccionado.
- f_i =Frecuencia absoluta del intervalo seleccionado.
- f_{i+1} =Frecuencia absoluta de la clase siguiente al intervalo seleccionado.
- f_{i-1} =Frecuencia absoluta de la clase anterior al intervalo seleccionado.

En la expresión:

$$M_o = L_i + a \cdot \left[\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})}\right]$$

En esta situación el intervalo con mayor frecuencia es $[60-70[$, por lo tanto, los datos serían:

- $a = 70 - 60 = 10$
- $L_i = 60$
- $f_i = 13$
- $f_{i+1} = 4$
- $f_{i-1} = 10$

Al sustituir en la expresión antes mencionada tenemos:

$$M_o = L_i + a \cdot \left[\frac{f_i - f_{i-1}}{(f_i - f_{i-1}) + (f_i - f_{i+1})}\right]$$
$$M_o = 60 + 10 \cdot \left[\frac{13-10}{(13-10)+(13-4)}\right]$$
$$M_o = 60 + 10 \cdot \frac{1}{4}$$
$$M_o = 60 + \frac{5}{2}$$
$$M_o = 60 + \frac{5}{2}$$
$$M_o = \frac{125}{2}$$
$$M_o = 62,5$$

Por lo tanto, si queremos ser más precisos con el valor de la moda, esta sería de 62,5 minutos.

Desafío 1

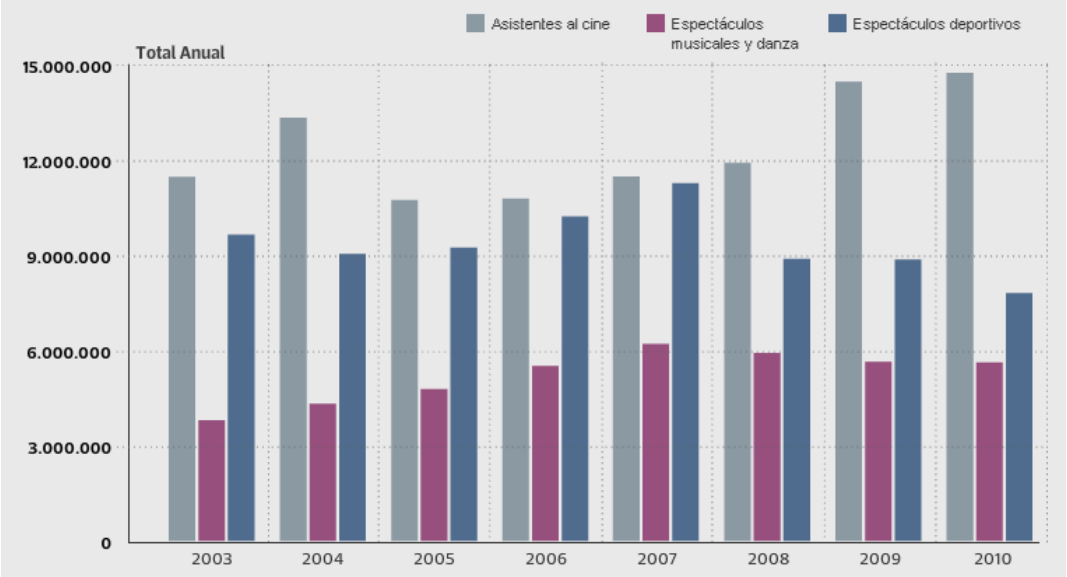


Escribir dos muestra de datos con 12 elementos cada una. La primera debe cumplir con que todas las medidas de tendencia central deben estar dentro de los valores de la muestra y la segunda debe cumplir con que todas las medidas de tendencia central no deben estar dentro de los valores de la muestra. Respuesta

✍ *Ejercicios* 2

1. El Instituto Nacional de Estadística (INE) presentó los resultaods del censo desde al año 2003 al año 2010 sobre las actividades que realizaban los chilenos en su tiempo libre:

Años	Asistentes al cine	Espectáculos musicales y danza	Espectáculos deportivos
2003	11.444.907	3.802.105	9.632.742
2004	13.301.754	4.312.663	9.025.897
2005	10.722.860	4.780.771	9.227.115
2006	10.763.165	5.505.433	10.202.285
2007	11.455.550	6.198.478	11.250.969
2008	11.886.801	5.915.303	8.873.371
2009	14.442.596	5.634.726	8.850.243
2010	14.714.031	5.614.614	7.796.126



a) Calcular las tres medidas de tendencia central para cada serie de datos.

b) ¿Qué comparaciones puedes obtener a partir de los resultados?

2. Una nutricionista decide estudiar el efecto que tienen dos tipos de dietas para bajar de peso durante 1 mes en sus pacientes. A continuación se muestran los resultados de los kilogramos bajados por dos grupos de pacientes sometidos unos a la dieta A y otros a la dieta B:

Dieta A

0,5	3,4	2,2	0,8	2,0	1,3	4,1	3,8	0,9
1,6	1,8	4,0	3,6	2,7	3,4	3,0	2,4	2,6

Dieta B

5,0	3,5	3,0	0,4	, 1,3	0,5	0,8	1,4	1,6
2,5	1,5	0,7	0,4	2,8	3,3	3,9	1,2	0,4

a) Calcular las tres medidas de tendencia central para cada serie de datos.

b) Ordenar la información en dos tabla de frecuencia con datos agrupados en clases o intervalos.

c) Calcular a partir de las tablas las medidas de tendencia central para cada serie de datos.

d) ¿Qué conclusiones puedes obtener a partir de los resultados del punto a) y c)?

Desafíos resueltos

✓ Desafío I:

1. **Muestra I:**

45,2	45,5	45,2	47,4	48,3	50,2
50,2	59	65,5	65,5	92,4	93,9

Las medidas de tendica central son:

- $\overline{X} = 59$
- $M_e = 50,2$
- $M_o = 45,2$

2. **Muestra II:**

1	2	3	4	5	6
7	8	9	10	11	12

Las medidas de tendica central son:

- $\overline{X} = 6,5$
- $M_e = 6,5$
- $M_o =$No existe

Volver

Bibliografía

[1] Manual de preparación PSU Matemática, *Quinta Edición*,
Oscar Tapía Rojas, Miguel Ormazábal Díaz-Muñoz, David López, Jorge Olivares Sepúlveda.

[2] Desarrollo del pensamiento matemático, Introducción a la Estadística, No 17, *Junio 2007*,
Martín Andonegui Zabala.

[3] Introducción a la Estadística, *Segunda Edición, 2007*,
Sheldom M.Ross.