Capítulo 8

MEDIDAS DE TENDENCIA CENTRAL

Sumario

Objetivos específicos

Al finalizar el estudio del capítulo, el estudiante será capaz de:

1. Definir cada una de las principales medidas de posición.
2. Calcular e interpretar el significado de las principales medidas de posición, tanto en datos agrupados como en datos no agrupados.
3. Seleccionar la medida de posición más adecuada para una determinada situación.

Resumen

En este capítulo se estudian las medidas de tendencia central, también llamadas "promedios". Para cada una se discute el concepto básico sobre el cual descansa, su definición y el procedimiento que se sigue para su cálculo, tanto en datos disponibles individualmente como agrupados. Se analizan también las características de las medidas y se dan algunas guías que permitan seleccionar la adecuada en una situación concreta.

8.1. LAS MEDIDAS DE TENDENCIA CENTRAL: IDEAS BÁSICAS

Cuando se trató el tema de las distribuciones de frecuencias, se hizo ver la importancia que tiene, para el análisis estadístico, el contar con elementos o valores descriptivos de tres características de los conjuntos de datos: a) la forma o patrón de distribución de los datos; b) su posición o tendencia central (centro de los datos) y c) la dispersión o variabilidad alrededor de los valores centrales. La distribución de frecuencias y su representación gráfica –por medio de histogramas o polígonos– ayudan, significativamente, a visualizar y conocer estos aspectos, en especial el primero. Es necesario, sin embargo, tanto para el análisis e interpretación del conjunto de datos, como para realizar comparaciones entre varios conjuntos de valores, el tener medidas que resuman o condensen en una forma sucinta, las características del conjunto en cuanto al punto donde se centran los datos (tendencia central) y la variabilidad o dispersión respecto a él. En este capítulo se tratarán las medidas de tendencia central y en el próximo, las de variabilidad.

El propósito básico de las medidas de tendencia central es resumir, en un solo número, el *centro de los datos* o punto central de localización de la distribución.[1] Puesto que el centro de una distribución puede ser definido en varias formas, existen diferentes medidas de tendencia central: la más conocida es la media aritmética, la cual corresponde al promedio simple de los valores. Otras dos muy importantes son la moda, que concierne al valor más frecuente o más probable, y la mediana, esta es el valor central que divide los datos en la mitad superior y la mitad inferior de ellos. También existen unas medidas de importancia menor como la media geométrica y la media armónica.

1. Cuando la distribución es particular, no es acampanada, las medidas de tendencia central (o algunas de ellas) no serán necesariamente valores centrales.

Es importante señalar que otro término utilizado, en el lenguaje corriente, para hacer referencia a las medidas de tendencia central, es el de *promedios*. Los estudiantes y profesores, por ejemplo, están acostumbrados al concepto de "nota promedio" y saben que al final de los cursos una actividad muy importante es el "cálculo de los promedios" y su empleo para decidir si un estudiante aprueba el curso, debe presentarse a un examen especial o repetirlo. Igualmente, en la vida diaria se habla de peso, salario, temperatura promedio. Pero también del individuo promedio o medio, para hacer referencia a una persona que refleja –dentro de ciertos límites– las características consideradas más corrientes o típicas de los miembros de una población o comunidad.

Mientras que la "nota promedio" puede obtenerse con operaciones aritméticas más o menos sencillas, el "individuo promedio" no es fácil de identificar en esa forma; esto porque, como se mencionó, es que hay diferentes formas de caracterizar el centro de una distribución, y por ende, se tienen diversos tipos de conceptos incluidos dentro de la noción corriente de promedio, aunque todos tiendan a dar una idea de lo típico o característico del fenómeno en consideración.

Como el lenguaje estadístico pretende ser más preciso que el usado popularmente, el término "promedio" se emplea con una significación más concreta, como una forma alternativa de designar la *media aritmética*.[2] Los demás conceptos involucrados en la noción corriente de promedio corresponden, asimismo, a otras medidas de posición correctamente definidas como la moda y la mediana.

En resumen, la tendencia central o "centro" de un conjunto de datos puede expresarse en varias formas o considerando diferentes dimensiones, por ello la estadística utiliza varias medidas de tendencia central:

a) Media aritmética o promedio
b) Mediana
c) Moda
d) Media geométrica
e) Media armónica

Las tres primeras son las de mayor importancia y, dentro de ellas, la *media aritmética* es la de utilización más frecuente y generalizada. La *media geométrica* y *la armónica* tienen un uso bastante restringido.

2. En lo que sigue se utilizará el término promedio únicamente para hacer referencia a la media aritmética.

8.2. SÍMBOLO DE SUMATORIA

Para el tratamiento de las medidas de tendencia central y de variabilidad y, en general, para definir las medidas estadísticas, deducir las fórmulas y examinar sus propiedades, resulta muy provechoso establecer un sistema de símbolos, es decir, una notación.

Se utilizará, generalmente, la letra x para indicar la variable en consideración: peso, ingreso, nota en un curso, número de bombillos defectuosos, producción por hora, resultado de una medición, entre otras. Un subíndice i que toma valores enteros positivos, servirá para indicar el elemento i del grupo estudiado; en consecuencia, x_i representa el valor particular o específico de la variable para ese elemento i. Por ejemplo, si se han pesado 6 estudiantes y se han obtenido los siguientes valores (en kilos): 55, 64, 53, 79, 64, 68, pueden indicarse simbólicamente en la siguiente forma:

$$x_1 = 55,\ x_2 = 64,\ x_3 = 53,\ x_4 = 79,\ x_5 = 64,\ x_6 = 68,$$

donde x_1 representa el peso del estudiante número 1; x_2, el peso del estudiante número 2; etc. Note que el estudiante número 2 y el 5 pesan ambos 64 kg; sin embargo, tienen diferentes subíndices porque se trata de dos individuos diferentes. El subíndice, por lo tanto, sirve para distinguir una observación de otra; también puede ser utilizado para indicar orden de selección, de magnitud o algún otro que interese.

El cálculo de las medidas estadísticas requiere, con gran frecuencia, sumar cierto número de valores de una característica. Esta operación se indica con el símbolo $\sum$ (**sigma**) o de *sumatoria*. Por ejemplo, si se tienen n valores de una característica

$$x_1, x_2, x_3, \ldots\ldots, x_{n-1}, x_n;$$

y se quiere indicar la suma de ellos, esta se expresa de la siguiente forma:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + x_4 + \ldots\ldots\ldots + x_{n-1} + x_n.$$

El símbolo $\sum_{i=1}^{n} x_i$ se lee "sumatoria de equis sub-i desde i igual a 1 hasta i igual a n". El subíndice i toma valores enteros consecutivos desde 1 hasta n. Los números o letras que aparecen debajo y encima de $\sum$ indican la extensión de la sumatoria.

Si $\sum_{i=1}^{n}$ aparece delante de una variable o expresión con subíndice i, esto indica que deben sumarse todos los valores particulares de la variable o expresión originados al darle a i los valores enteros 1, 2,... hasta n, inclusive.

Suponga que interesa la variable x igual al peso en kilogramos, y que se tiene la muestra de $n = 6$ estudiantes antes mencionada. En este caso, $\sum_{i=1}^{6} x_i$ indica que se deben sumar los 6 valores de x_i que van de x_1 a x_6, inclusive. Entonces:

$$\sum_{i=1}^{6} x_i = x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 55 + 64 + 53 + 79 + 64 + 68 = 383.$$

La $\sum$, por lo tanto, es un símbolo que pide ejecutar la operación de suma de la variable de interés, haciendo variar el subíndice de acuerdo con los límites indicados. Así, por ejemplo, si se tuvieran los 6 pesos antes mencionados y se pidiera calcular $\sum_{i=1}^{6} x_i^2$, dicha expresión ordena elevar al cuadrado los pesos de los 6 estudiantes y luego sumarlos, de la forma siguiente:

$$\sum_{i=1}^{6} x_i^2 = x_1^2 + x_2^2 + ... + x_6^2 = 55^2 + 64^2 + ... + 68^2 = 24\,891.$$

A continuación, se presentan varias ilustraciones del uso del símbolo de sumatoria:

1. $\sum_{i=1}^{k} x_i f_i = x_1 f_1 + x_2 f_2 + x_3 f_3 + ... + x_k f_k.$
2. $\sum_{i=1}^{8} i = 1 + 2 + 3 + ... + 8 = 36.$

   Note que la variable x es a la vez la variable de la sumatoria.
3. $\sum_{j=1}^{6} f_j(x_j + z_j) = f_1(x_1 + z_1) + f_2(x_2 + z_2) + ... + f_6(x_6 + z_6).$
4. $\sum_{i=4}^{10} y_i^2 = y_4^2 + y_5^2 + y_6^2 + y_7^2 + y_8^2 + y_9^2 + y_{10}^2.$

   Note que se empezó en $i = 4$ porque así lo indica el límite inferior de la sumatoria.
5. $\left(\sum y_i\right)^2 = (y_1 + y_2 + y_3 + y_4 + ... + y_i + ... + y_n)^2.$

Una práctica común es omitir el subíndice y los límites de las sumatorias siempre que su ausencia no implique el peligro de confusiones o ambigüedades.

El símbolo de sumatoria tiene algunas propiedades de mucha utilidad, las cuales se resumen en el cuadro 8.1.

Cuadro 8.1

PROPIEDADES DEL SÍMBOLO DE SUMATORIA $\sum$

PROPIEDAD 1	La sumatoria del producto de una constante por una variable es igual al producto de la constante por la sumatoria de la variable	$\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i$
PROPIEDAD 2	La sumatoria de la suma algebraica de dos o más variables es igual a la suma algebraica de las sumatorias de las variables	$\sum_{i=1}^{n}(x_i + y_i - z_i) = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} z_i$
PROPIEDAD 3	La sumatoria de una constante, tomada de 1 a n, es igual a n veces la constante	$\sum_{i=1}^{n} a = na$

Ejercicios ilustrativos

1. $\sum_{i=1}^{m}(x_i + a - bz_i) = \sum_{i=1}^{m} x_i + \sum_{i=1}^{m} a - \sum_{i=1}^{m} bz_i$

$$= \sum_{i=1}^{m} x_i + ma - b\sum_{i=1}^{m} z_i.$$

2. $\sum_{j=3}^{k}(ax_j - b) = \sum_{j=3}^{k} b$

$$= a\sum_{j=3}^{k} x_j - b(k-2).$$

8.3. MODA, MEDIANA Y MEDIA ARITMÉTICA EN DATOS NO AGRUPADOS

Los datos con los que se realiza el análisis estadístico pueden estar disponibles, individualmente o agrupados, en una distribución de frecuencias. Los procedimientos de cálculo de las medidas estadísticas varían un poco dependiendo de cuál de estas dos situaciones se trate. Se presentará primero aquella en la que los datos no están agrupados y, posteriormente, se hará referencia a cómo debe procederse si lo estuvieran. En ambos casos, se discutirá la definición y los procedimientos de cálculo de las tres medidas de tendencia central más importantes; luego se hará una referencia a la media geométrica y a la armónica.

8.3.1. La moda (Mo)

La moda de un conjunto de datos se define como el valor más frecuente, el que más se repite, indica el punto donde la variable ocurre con mayor densidad. Para calcularla, en datos no agrupados, se obtiene la frecuencia de los valores e identificar aquel que exhibe la mayor frecuencia. En la serie incluida seguidamente se puede observar que la moda es 21, al ser el valor al cual corresponde la mayor frecuencia.

$$14, 15, 17, 17, 21, 21, 21, 21, 33, 36, 40.$$

La moda es una medida muy natural para describir un conjunto de datos, es muy simple de calcular y quizá la más fácil de interpretar de las medidas de tendencia central.

Su concepto se adquiere fácilmente: es el sueldo más común, el peso más corriente, la edad más frecuente, el número de hijos más usual. La moda, por lo general, es lo que tienen en mente las personas cuando hablan de "promedios". Se acerca mucho, por lo tanto, al concepto o idea que se maneja cuando se habla del costarricense promedio, del votante típico, del ama de casa común, la familia promedio. A esto contribuye el hecho de que en valores no agrupados, su valor siempre corresponde a un valor real observado, mientras que la mediana en especial la medida aritmética pueden asumir valores con fracciones que no corresponden a los observados.

Tiene la ventaja de que no se ve afectada por la presencia de valores extremos, altos o bajos y, por otra parte, puede ser aplicada a variables de todos los niveles de medición, siendo especialmente útil para el caso de las nominales.

Ejemplo 1

En una encuesta realizada en setiembre del 2009, la cual abarcó una muestra de 350 costarricenses adultos, residentes en la Región Metropolitana de San José, se preguntó cuál era el centro comercial *(mall)* que más le gustaba visitar. Las respuestas dieron la siguiente distribución, donde se aprecia que la moda corresponde a Multiplaza de Escazú, que es, por lo tanto el *mall* con la mayor preferencia de los entrevistados.

Cuadro 8.2

CENTRO COMERCIAL DE PREFERENCIA,
ENCUESTA REALIZADA EN SETIEMBRE DEL 2009

CENTRO COMERCIAL	% LO MENCIONÓ	MEDIDA
Multiplaza de Escazú	25,7	Moda
Mall San Pedro	23,4	
Paseo de las Flores, Heredia	15,2	
Terramall, Tres Ríos	14,9	
Real Cariari, Belén	5,6	
Multiplaza del Este	5,0	
Mall Internacional, Alajuela	3,2	
Otro	7,0	
TOTAL	100,0	

La moda, aunque tiene algunas buenas características como su sencillez conceptual y la simplicidad de su cálculo, presenta desventajas importantes, por las cuales se utiliza menos que la media aritmética y la mediana. Su principal limitación está en el hecho de que requiere un número suficiente de observaciones para manifestarse claramente. Además, en ciertos casos puede no existir, no estar definida; también es posible que exista pero no ser única, en el sentido de que otro valor de la distribución tenga aún una frecuencia elevada. En este último caso, la serie de datos tiene más de un valor modal y no es claro el criterio para decidir cuál debe usarse como moda. Por ejemplo, en la siguiente serie la moda no está definida, al no repetirse ninguno de los datos:

152, 178, 160, 148, 165, 155, 164.

Y en esta otra

$$50, 55, 55, 55, 62, 73, 73, 73, 80,$$

se presentan dos modas, ya que los valores 55 y 73 se repiten tres veces.

Igualmente, puede darse el caso de distribuciones como la siguiente, donde se define una moda "mayor" (igual a 2, con 7 frecuencias) y una "menor" (igual a 5, con 3 frecuencias).

$$1, 2, 2, 2, 2, 2, 2, 3, 3, 4, 5, 5, 5, 7, 9.$$

En estos casos, se dice que la distribución es bimodal y se tiene una situación en que la moda es muy difícil de interpretar, en consecuencia, carece de utilidad como medida de tendencia central.

Existe otra desventaja originada en su sensibilidad por la forma en que se definen los intervalos cuando los datos se agrupan en una distribución de frecuencias.

◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇◇

8.3.2. La mediana (Me)

La mediana es una típica medida de posición y se define como el valor central de una serie de datos ordenados o, más precisamente, como un valor tal que no más de la mitad de las observaciones son menores que ella y no más de la mitad mayores. De acuerdo con esto, en un grupo de datos ordenados la mitad precede a la mediana y la otra mitad lo superan.

Cuando se tienen datos sin agrupar y se desea calcular la mediana es necesario, en primer lugar, ordenarlos de acuerdo con su magnitud. Luego, se determina el valor central de la serie y esa es la mediana. Si el número de datos es par, existirán dos valores centrales, entonces se obtiene haciendo el promedio de ellos. El cálculo se facilita recordando que el valor central de una serie de n datos es el $\frac{n+1}{2}$ término de la serie.

Considérense los datos siguientes, en que n es impar ($n = 7$):

$$6, 8, 8, \mathbf{10}, 12, 19, 23$$

$$\frac{n+1}{2} = \frac{7+1}{2} = 4.$$

Me = 10 (valor que corresponde al 4° término).

Ahora considere un caso en que n es par ($n = 8$).

$$3, 4, 4, \mathbf{5}, \mathbf{16}, 19, 25, 30.$$

$$\frac{n+1}{2} = \frac{8+1}{2} = 4,5\,.$$

$$Me = \frac{5+16}{2} = 10,5 \text{(valor que corresponde al 5° término).}$$

Note que, en este caso, la mediana está situada entre el 4° y el 5° término, por ello su valor se estima promediando los valores correspondientes a esos términos, o sea, los valores 5 y 16.

La mediana tiene una ventaja importante: siempre puede calcularse y el valor obtenido está bien definido (lo cual no es cierto para la moda). Además, al ser una medida posicional, tiene la importante ventaja de que no es afectada por valores extremos, como sí lo es la media aritmética.

Tiene también una propiedad interesante, muy útil en ciertas oportunidades, y es que la suma de los valores absolutos de las desviaciones de los valores, con respecto a la mediana, es menor que las desviaciones con respecto a cualquier otro valor.

Sus principales limitaciones son: a) es un valor calculado y que no siempre coincide con el de un dato observado; y b) por su naturaleza no puede ser usado en muchos procedimientos estadísticos, cosa que si es posible con la media aritmética.

8.3.3. La media aritmética ($\overline{X}$)

La media aritmética es la medida de tendencia central más usada y conocida. Corrientemente se le llama "promedio" y, de ahora en adelante, cuando se diga promedio, se hace referencia a la media aritmética. También es frecuente llamarla "equis barra". En la práctica, dependiendo de la forma como estén disponibles los datos y los objetivos, puede calcularse como una *media aritmética simple* o como una *media aritmética ponderada*.

8.3.4. La media aritmética simple

La media aritmética de un conjunto de valores es *el resultado que se obtiene al dividir la suma de esos valores entre el número de ellos.*

$$\text{Medida aritmética} = \frac{\text{Suma de los valores}}{\text{Número de valores}}$$

En general, si la característica considerada es x y se tienen n valores:

$$\bar{X} = \frac{1}{n}\sum x_i = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Por ejemplo, si en una sala hay 12 personas cuyas edades son:

20, 20, 22, 20, 30, 25, 25, 18, 20, 18, 22, 36.

La edad promedio de esas personas es:

$$\bar{X} = \frac{20 + 20 + 22 + ... + 22 + 36}{12} = \frac{276}{12} = 23 \text{ años.}$$

La media simple de las edades es, por lo tanto, 23 años.

8.3.5. Media aritmética ponderada

En el cálculo de la media aritmética simple, cada una de las observaciones recibe el mismo peso o ponderación. En ciertas situaciones, sin embargo, interesa promediar un conjunto de valores, dándole a cada uno de ellos un peso diferente, en forma desigual. En otras palabras, se desea calcular una media aritmética ponderada.

Supóngase que se tienen n valores de x: $x_1, x_2, x_3, \ldots, x_i, \ldots, x_n$, los cuales se desea promediar ponderándolos con los pesos $w_1, w_2, w_3, \ldots, w_i, \ldots, w_n$. En este caso, la fórmula para el cálculo de la media aritmética ponderada es la siguiente:

$$\bar{X}_w = \frac{\sum x_i w_i}{\sum w_i}.$$

Donde x_i representa el valor i de la variable de interés, w_i la ponderación que corresponde a ese valor x_i y x_w es el valor de la media ponderada. Cada valor se multiplica por la ponderación w_i que le corresponde y la suma de los productos se divide entre la suma de las ponderaciones.

Ejemplo 2

Un pequeño empresario compró 1000 dólares en enero del 2010, cuando el precio de venta era de 570 colones; 1000 en febrero a un precio de 553 colones y 4000 más, en marzo, a 536 colones. Supóngase ahora que, para ciertos propósitos, el empresario desea calcular el precio promedio que pagó por esas compras de dólares del primer trimestre del año 2010.

Cuadro 8.3

PRECIO PROMEDIO DE COMPRAR DE DÓLARES EN EL 2010

MES	PRECIO COMPRA	MONTO COMPRADO	MONTO PAGADO
	x_1	w_1	$x_1 w_1$
Enero	570	1000	570 000
Febrero	553	1000	553 000
Marzo	536	4000	2 144 000
TOTAL		6000	3 267 000

$$\bar{X}_w = \frac{\sum x_i w_i}{\sum w_i} = \frac{570 \cdot 1000 + 553 \cdot 1000 + 536 \cdot 4000}{6000}$$

$$= \frac{3\,267\,000}{6000} = 554,50.$$

El pequeño empresario compró los $6000 a un precio promedio de 544,50 colones. Note dos cosas:

a) El promedio simple de los precios pagados (553) no arroja el valor correcto, porque el monto de dólares comprados no fue igual cada mes. Al ser diferentes, los precios deben ser ponderados por el monto comprado en cada oportunidad.

b) Si los montos comprados cada mes hubieran sido iguales ($2000), el promedio simple de los precios daría el valor correcto, porque todos los precios tendrían igual peso, igual ponderación.

Ejemplo 3

En los datos de edades de 12 personas utilizados para ilustrar el cálculo de la media simple, se nota que varias se repiten. Si se agrupan, se obtiene la distribución indicada abajo. Es evidente que si se multiplica cada edad por el número de personas que la tienen y luego se suman esos productos, el total será 276 y se llegará al mismo resultado que si se hubieran sumado todos los valores uno por uno.

Cuadro 8.4

DATOS DE LAS EDADES DE 12 PERSONAS

Edad X	Nº de personas F	xf
18	2	36
20	4	80
22	2	44
25	2	50
30	1	30
36	1	36
TOTAL	12	276

$$\bar{X}_w = \frac{2 \cdot 18 + 4 \cdot 20 + ... + 1 \cdot 36}{2 + 4 + ... + 1} = \frac{276}{12} = 23 \text{ años}$$

En esta forma de cálculo del promedio, lo que se ha hecho es ponderar cada edad diferente *(x)* por la frecuencia de aparición en el grupo de datos *(f)*, sumar estos productos y el resultado de la suma dividirlo por la suma de las frecuencias (ponderaciones), es decir, se ha calculado un promedio ponderado.

Los promedios ponderados son muy corrientes en los colegios y universidades, donde los estudiantes, muchas veces sin saberlo, los calculan. Esto sucede porque es usual darle "pesos" o ponderaciones diferentes a las notas de los exámenes parciales, "quices" y trabajos prácticos que se combinan para calcular la nota de presentación, la cual debe ser integrada con la nota del examen final para obtener la calificación final.

Ejemplo 4

Un profesor realiza tres exámenes durante el semestre con el número de preguntas y la duración que abajo se indican. En ellos, la estudiante Laura Solís Guevara obtuvo, respectivamente, 88,62 y 63 puntos.

Cuadro 8.4

RESULTADO DE EXÁMENES PARCIALES

Examen parcial	N° preguntas	Duración (minutos)	Nota de Laura Solís Guevara
I	9	50	88
II	5	80	62
II	6	120	63

a) Si se calcula la nota del curso en la forma usual, dándole a cada examen igual peso, ¿qué nota obtiene Laura Solís Guevara?

$$\bar{X} = \frac{88 + 62 + 63}{3} = \frac{213}{3} = 71.$$

b) ¿Y si se pondera cada nota por el número de preguntas en que se basa?

$$\bar{X} = \frac{9 \cdot 88 + 5 \cdot 62 + 6 \cdot 63}{9 + 5 + 6} = \frac{1480}{20} = 74.$$

Como puede observarse, la nota resulta más alta que la obtenida utilizando el promedio simple empleado en a).

c) ¿Y si se pondera por la duración del examen?

$$\bar{X} = \frac{50 \cdot 88 + 80 \cdot 62 + 120 \cdot 63}{50 + 80 + 120} = \frac{16920}{250} = 67,68 \approx 68.$$

Puede observarse que la nota promedio da diferentes resultados, según el criterio empleado para calcularla, siendo la más alta aquella que se obtiene usando como ponderación el número de preguntas. Esto sucede debido a que el examen parcial con el cual Laura obtuvo la nota más alta (88) fue precisamente aquel donde el número de preguntas fue más elevado. Igualmente, el hecho de que su nota más baja se dio en el examen de mayor duración lleva a que el promedio ponderado que utiliza la duración como criterio arroja la nota más baja.

Ahora bien, ¿cuál es el promedio "correcto"? En realidad, no hay uno "correcto". Se supone que, antes de iniciarse el curso, se ha fijado el criterio para calcular la nota –posiblemente el de que cada examen tendrá el mismo peso– y ese producirá la nota oficial del curso. Pero suponga, por un momento, que por alguna circunstancia ese criterio no hubiera sido fijado de manera oportuna, ¿cuál sería el método más favorable a la estudiante Laura Solís? Es evidente, el que utiliza como ponderación el número de preguntas del examen, pues ese le da la nota más alta (74 puntos). Pero ¿cuál convendría más a Óscar Alberto Castillo, quien obtuvo respectivamente 60, 68 y 76 en los exámenes parciales? Si se hace el cálculo, se encuentra que la nota del curso resulta 68 (calculada en la forma usual), 66,8 si se pondera por el número de preguntas y 70,2 si el criterio empleado es la duración de la prueba. Obviamente, a Óscar Alberto le favorece más el último procedimiento de cálculo.

No es difícil imaginar la discusión que se plantearía entre ambos estudiantes y los argumentos para "demostrar" que el método que les conviene es el más "justo" o el "correcto". Por este motivo, los criterios de calificación y de cálculo de los promedios se fijan normalmente antes del inicio del curso y de la realización de los exámenes.

8.3.6. Media aritmética global de varios conjuntos de datos

Cuando se tienen las medias aritméticas para varios conjuntos de datos, y se conoce también el número de observaciones que los componen, la obtención de la media aritmética para el total de los datos implica el cálculo de un promedio ponderado. Por ejemplo:

Conjunto 1	**Conjunto 2**	**Conjunto 3**
N_1	N_2	N_3
X_1	X_2	X_3

$$\text{Medida general} = \bar{X} = \frac{N_1X_1 + N_2X_2 + N_3X_3}{N_1 + N_2 + N_3}.$$

Como puede apreciarse, la media general no es ni más ni menos que la ponderada de las medias de los conjuntos, donde las ponderaciones son el número de elementos de cada uno de los conjuntos.

8.4. PROPIEDADES DE LA MEDIA ARTIMÉTICA

Por su utilidad práctica y analítica, es importante conocer algunas propiedades matemáticas que tiene la media aritmética. Se enuncian e ilustran seguidamente:

PROPIEDAD 1. Si se multiplica la media por el número de observaciones, se obtiene la suma de las observaciones.

$$n\overline{x} = \sum_{i=1}^{n} x_i .$$

PROPIEDAD 2. Si a cada una de las observaciones se le resta la media, y luego se suman esas desviaciones –o diferencias– la suma resulta igual a 0.

$$\sum^{n}(x_i - \overline{x}) = 0 .$$

PROPIEDAD 3. Si se suma (o se resta) una constante b a cada una de las observaciones, el promedio aritmético queda aumentado (o disminuido) en esa constante b.

$$y_i = x_i + b, \text{ entonces } \overline{y} = \overline{x} + b.$$

$$y_i = x_i - b, \text{ entonces } \overline{y} = \overline{x} - b .$$

PROPIEDAD 4. Si se multiplica (o se divide) cada una de las observaciones por una constante b, el promedio aritmético queda multiplicado (o dividido) por esa constante b.

$$y_i = bx_i, \text{ entonces } \overline{y} = b\overline{x}.$$

$$y_i = \frac{x_i}{b}, \text{ entonces } \overline{y} = \frac{\overline{y}}{b}.$$

Ejemplo 5 (Propiedades 3 y 4)

Un profesor realiza un examen y después de calificarlo calcula el promedio, obteniendo $\overline{x} = 68$. Ante una gestión de los alumnos, decide aumentarle 5 puntos a cada uno. ¿Cuál será el nuevo promedio?

$$\overline{y} = \overline{x} + b = 68 + 5 = 73.$$

Al comentar el asunto con el director de cátedra, este critica la decisión aduciendo que los alumnos deben ser favorecidos en proporción al resultado obtenido, y lo convence de que mejor les aumente en un 10% la nota. ¿Cuál será el nuevo promedio con esta decisión?

$$\overline{y} = b\overline{x} = 1,10 \cdot 68 = 74,8.$$

◇◇

Ejemplo 6 (Propiedad 2)

En el ejemplo 3, referente a la edad de 12 personas, se obtuvo $\overline{x} = 23$ años. Compruebe, en ese grupo de datos, que la suma de las desviaciones con respecto al promedio es igual a 0

$$\sum_{i=1}^{n}(x_i - \overline{x}) = 0.$$

$20 - 23 = -3$	$20 - 23 = -3$	$22 - 23 = -1$	$20 - 23 = -3$
$30 - 23 = 7$	$25 - 23 = 2$	$25 - 23 = 2$	$19 - 23 = -5$
$20 - 23 = -3$	$18 - 23 = -5$	$22 - 23 = -1$	$36 - 23 = 13$

Suma de desviaciones:

$$(-3) + (-3) + (-1) + (-3) + (7) + (2) + (2) + (-5) + (-3) + (-1) + (13) = 0.$$

◇◇

Ejemplo 7 (Propiedad 1)

Un estudio por muestreo, realizado en un país hace un año, mostró que el consumo anual de camarones por familia se situaba en alrededor de 1,5 kilos. Se desea estimar el consumo para el presente año y se sabe que el número de familias es de aproximadamente 2 millones ¿En cuánto estimaría dicho consumo?

$$N\overline{x} = 2000000 \cdot 1,5 = 3000000\,\text{kg}.$$

◇◇

8.5. EL CÁLCULO DE LAS MEDIDAS DE POSICIÓN EN DATOS AGRUPADOS

En muchos casos, los datos disponibles para realizar un análisis estadístico están agrupados en una distribución de frecuencias. Bajo estas condiciones, no es posible utilizar directamente las fórmulas y procedimientos vistos antes, sino que deben derivarse fórmulas específicas que tomen en cuenta las características de las distribuciones de frecuencias. A continuación, se presentan las fórmulas para el cálculo de la moda, mediana y promedio aritmético cuando los datos están agrupados en una distribución de frecuencias.

8.5.1. La moda (Mo)

La moda fue definida como el valor del conjunto de datos que se repite con más frecuencia, como el más común. Esta definición, sin embargo, no es aplicable directamente cuando se tienen los datos agrupados en clases. En este caso, lo más conveniente es ubicar la "clase modal", o sea, aquella clase de la distribución donde hay *mayor densidad de frecuencias por unidad de intervalo* y luego, utilizando algunas de las fórmulas propuestas, estimar el valor de la moda.

A continuación, se presenta una de las fórmulas de uso más difundido:

$$M_o = L_i + \frac{d_1}{d_1 + d_2} \cdot c\,.$$

Donde:

L_i = Límite inferior real de la clase modal.

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.

c = Intervalo de la clase modal.

Para ilustrar el uso de la fórmula, considérese el ejemplo del peso en kilogramos de los 60 alumnos hombres de dos colegios privados que se presentó en el capítulo 7.

Cuadro 8.5
PESO EN KILOGRAMOS DE 60 ALUMNOS

Peso en kilogramos	Número de alumnos (*f*)
44,5 – 49,5	1
49,5 – 54,5	6
54,5 – 59,5	8
59,5 – 64,5	19
64,5 – 69,5	9
69,5 – 74,5	6
74,5 – 79,5	5
79,5 – 84,5	3
84,5 – 89,5	3
TOTAL	60

La clase modal es 59,5 – 64,5 cuya frecuencia es 19, la más elevada.

Aplicando la fórmula:

$$\begin{aligned} M_o &= 59,5 + \frac{11}{11+10} \cdot 5 \\ &= 59,5 + \frac{11}{21} \cdot 5 \\ &= 59,5 + 2,62 = 62,12. \end{aligned}$$

El procedimiento arroja un valor modal de 62,1 kilos.

Cuando la distribución de frecuencias tenga intervalos desiguales, deberán considerarse las "frecuencias ajustadas", tanto al identificar la clase modal como al aplicar la fórmula.[3]

8.5.2. La mediana (Me)

En una distribución continua, la ordenada correspondiente a la mediana divide el área bajo la curva en dos partes iguales. Esta propiedad se toma como base para definir la mediana (Me), en datos agrupados. Así, la Mediana es el valor Me –de la característica x considerada– al cual corresponde la frecuencia acumulada $\frac{n}{2}$. Para calcularla, se recurre

3. El procedimiento de ajuste de las frecuencias fue tratado en el capítulo 7, "Distribución de Frecuencias", en la sección "Representación de distribuciones con intervalos desiguales".

entonces a la fórmula siguiente, que es una simple interpolación lineal[4] del valor de Me dentro de los límites de la clase donde se encuentra la mediana.

$$M_e = L_i + \frac{\frac{n}{2} - F_a}{f_i} \cdot c.$$

Donde:

$n=$ Número total de observaciones o suma de las frecuencias absolutas.

$L_i=$ Límite inferior real de la clase donde está la Mediana, o sea, donde se alcanza la frecuencia acumulada $\frac{n}{2}$.

$f_i=$ Frecuencia absoluta de la clase donde está la mediana.

$F_a=$ Frecuencia acumulada "menos de" de la clase anterior a la clase donde está la mediana.

$c=$ Intervalo de la clase donde está la mediana.

Para ilustrar el cálculo, considere de nuevo la distribución de frecuencias correspondiente al peso de los 60 alumnos de dos colegios privados.

Cuadro 8.6

DISTRIBUCIÓN DE 60 ALUMNOS HOMBRES DE DOS COLEGIOS DE ACUERDO CON SU PESO EN KILOGRAMOS

Clases	Frecuencia Absoluta (fi)	Frecuencia Acumulada (Fi)
44,5 - 49,5	1	1
49,5 - 54,5	6	7
54,5 - 59,5	8	15
59,5 - 64,5	19	34
64,5 - 69,5	9	43
69,5 - 74,5	6	49
74,5 - 79,5	5	54
79,5 - 84,5	3	57
84,5 - 89,5	3	60
TOTAL	60	

4. La interpolación consiste en un procedimiento para determinar un valor intermedio entre dos valores conocidos. La interpolación lineal es muy usada y supone que x varía linealmente dentro del intervalo; en ciertos casos, sin embargo, puede suponerse otro modelo de variación.

Proceso de cálculo:

1. Se obtienen las frecuencias acumuladas "menos de" (última columna del cuadro anterior).
2. Se calcula el valor de $\frac{n}{2} = \frac{60}{2} = 30$.
3. Se determina la clase donde está la mediana, es decir donde se alcanza la acumulada $\frac{n}{2}$. La clase es 59,5 – 64,5 , ya que hasta 59,5 hay acumuladas 15 observaciones y hasta 64,5 hay 34, y la mediana debe ser mayor que 59,5 y menor que 64,5.
4. Se calcula el intervalo: $c = 64{,}5 - 59{,}5 = 5{,}0$.
5. Se aplica la fórmula para el cálculo de la mediana:

$$M_e = 59,5 + \frac{30 - 15}{19} \cdot 5 = 59,5 + \frac{75}{19} = 63,45.$$

El resultado indica una $M_e = 63,45$, la cual dice que un 50% de los alumnos pesa menos de 63,45 kg y la otra mitad más de ese peso.

El cálculo de la mediana para una distribución con amplitudes desiguales de clase se hace utilizando el mismo procedimiento y la misma fórmula.

8.5.3. La media aritmética ($\overline{X}$)

En una distribución de frecuencias, se conoce el número de observaciones que hay dentro de una clase pero **no** se sabe el valor exacto de cada una de ellas; esta información se perdió al agrupar los datos. Surge entonces el problema de cómo obtener la suma de las observaciones que son requeridas para el cálculo del promedio; para hacerlo, se recurre a una hipótesis arbitraria, pero razonable: se supone que las observaciones dentro de una clase se distribuyen uniformemente dentro de la clase y que, por ello, el punto medio las representa adecuadamente. Como es evidente, esto equivale a suponer que todas son iguales al punto medio.

Hecho este supuesto, la suma de los valores de cada clase se obtiene multiplicando su punto medio (x_i) por su frecuencia (f_i), y la suma total haciéndola para todas las clases, dividiendo esa suma entre el total de frecuencias se obtiene la media aritmética. La fórmula resultante se incluye abajo y puede notarse que es simplemente el promedio de los puntos medios ponderados por las frecuencias.

$$\bar{X} = \frac{\sum x_i f_i}{\sum f_i} = \frac{\sum x_i f_i}{n},$$

donde:

$n = \sum f_i x_i$ representa el punto medio de la clase i.

f_i representa la frecuencia de la clase i.

Corrientemente, la media aritmética para una distribución de frecuencias difiere de la exacta, o sea, de la que se obtendría si se realizara el cálculo con los datos sin agrupar. La diferencia se debe al hecho de que los valores, dentro de las clases, no son todos iguales al punto medio y ni siquiera se distribuyen uniformemente dentro del intervalo de clase; por lo tanto, al suponerse que todos equivalen al punto medio, se está provocando una discrepancia con el valor exacto. Sin embargo, como la discrepancia no lleva el mismo sentido en todas las clases, sino que en unas se peca por defecto y en otras por exceso, globalmente se produce una compensación y, por consiguiente, la diferencia entre la media aritmética exacta –calculada con los datos individuales– y la obtenida con los datos agrupados resulta de poca importancia y puede despreciarse para efectos prácticos.

Seguidamente, se ilustra el cálculo de $\bar{X}$ para la distribución de frecuencias correspondiente al peso de los 60 alumnos hombres de dos colegios privados:

Cuadro 8.7

CÁLCULO DE $\bar{X}$ PARA LA DISTRIBUCIÓN DE FRECUENCIAS CORRESPONDIENTE AL PESO DE LOS 60 ALUMNOS HOMBRES DE DOS COLEGIOS PRIVADOS

PESO EN LIBRAS	PUNTOS MEDIOS (x_i)	FRECUENCIAS f_i	$x_i f_i$
44,5 – 49,5	47	1	47
49,5 – 54,5	52	6	312
54,5 – 59,5	57	8	456
59,5 – 64,5	62	19	1178
64,5 – 69,5	67	9	603
69,5 – 74,5	72	6	432
74,5 – 79,5	77	5	385
79,5 – 84,5	82	3	246
84,5 – 89,5	87	3	261
TOTAL		60	3920

$$\bar{X} = \frac{\sum xf}{n} = \frac{3920}{60} = 65,33.$$

Si se calcula el promedio con base en los datos originales, sin agrupar, según como aparecen en el capítulo 7, se obtiene, como fácilmente se puede comprobar:

$$\bar{X} = \frac{3896}{60} = 64,93.$$

Como puede apreciarse, la diferencia es pequeña, apenas alcanza un 0,4%.

8.6. USO DE LAS MEDIDAS DE TENDENCIA CENTRAL

El propósito fundamental de las medidas de tendencia central es caracterizar y representar el centro de un conjunto de datos, y cada una de las diferentes medidas propuestas lo hace en algún sentido.

La más usada es la media aritmética; sin embargo, en ciertas circunstancias, alguna de las otras medidas puede ser más adecuada, dependiendo de los propósitos que se tengan. Además, es conveniente hacer notar que, corrientemente, las medidas no compiten entre sí, sino que se complementan y permiten, en conjunto, una mejor descripción de los aspectos típicos del grupo de datos.

Una gran parte del problema de decidir qué medida usar, en una situación específica, desparece si se tiene una idea clara de cuál es el aspecto del conjunto de datos que se desea resumir y del efecto de los valores extremos sobre las medidas de tendencia central, bajos o altos.

En una distribución simétrica, como las dos que se ilustran abajo, la moda, la mediana y la media aritmética coinciden, es decir, valen lo mismo.[5] En este caso, cualquiera de esas medidas resulta igualmente adecuada para caracterizar el grupo de datos.

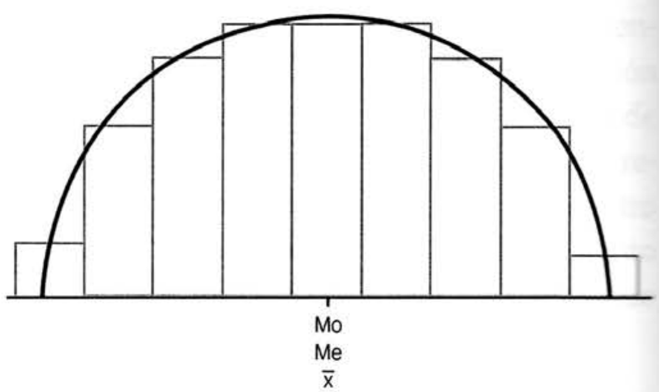


Figura 8.1 y figura 8.2. Ejemplo de distribuciones simétricas

5. Cuando la distribución es simétrica, pero en forma de U, solo la media aritmética y la mediana coinciden, ya que la moda no está definida.

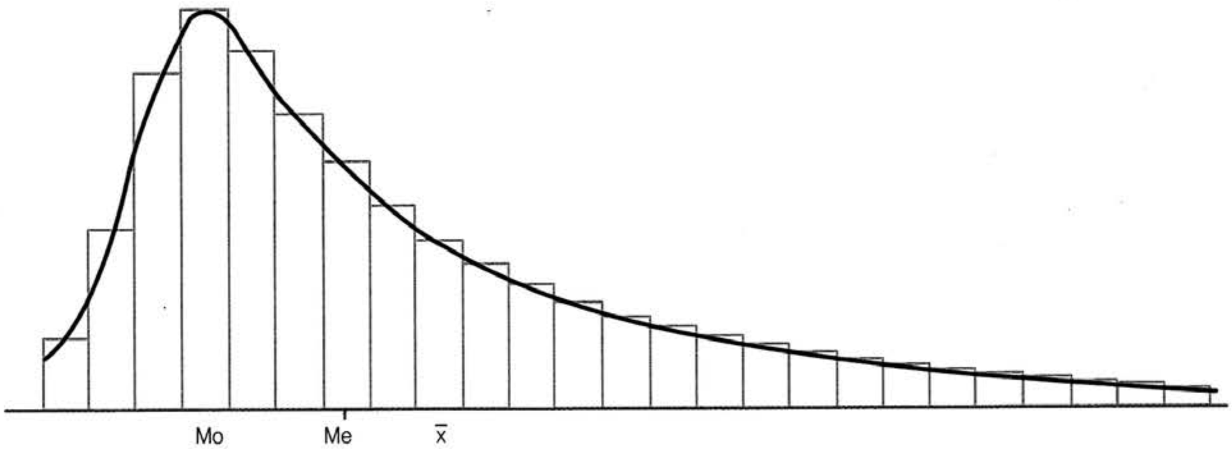


Figura 8.3. Ejemplo de distribución con asimetría positiva

En cambio, si la distribución es asimétrica, es decir, si una de sus colas es más extendida que la otra, la coincidencia entre las medidas de tendencia central desaparece y se puede presentar cualquiera de las dos situaciones que, a continuación, se ilustran:

a) *Asimetría positiva:* cola más larga hacia la derecha. La distribución tiene valores extremos altos.

$$\bar{X} > M_e > M_o.$$

b) *Asimetría negativa:* cola más larga hacia la izquierda. La distribución tiene valores extremos bajos.

$$\bar{X} < M_e < M_o.$$

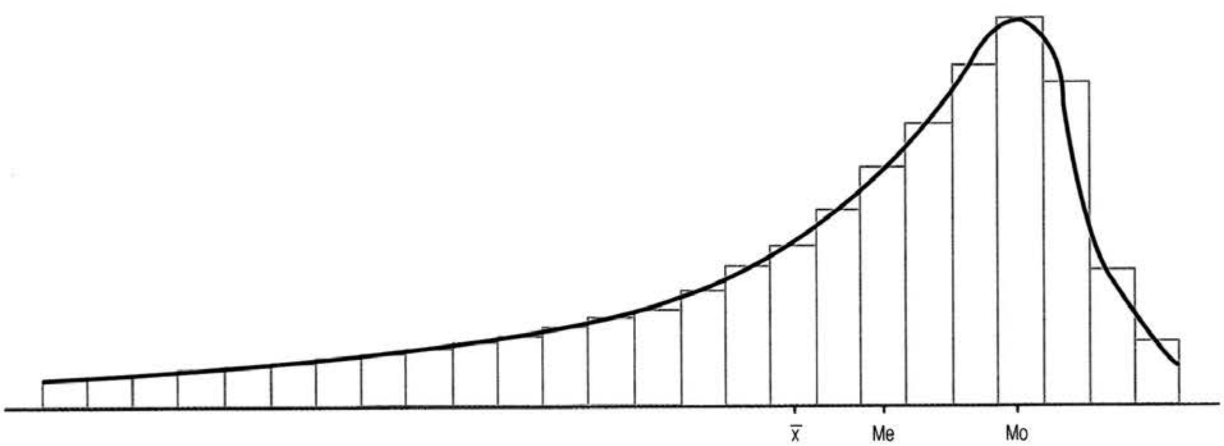


Figura 8.4. Ejemplo de distribución con asimetría negativa

Este comportamiento de las medidas de posición, en el caso de distribuciones asimétricas, caracterizado por un alejamiento, con respecto a la moda, de la mediana y en especial de la media aritmética, se debe a la influencia de los valores extremos.

Para ilustrar más claramente este punto, considérese el siguiente conjunto de datos correspondiente a una variable discreta distribuido simétricamente y su representación gráfica: 5, 6, 6, 7, 7, 7, 8, 8, 9.

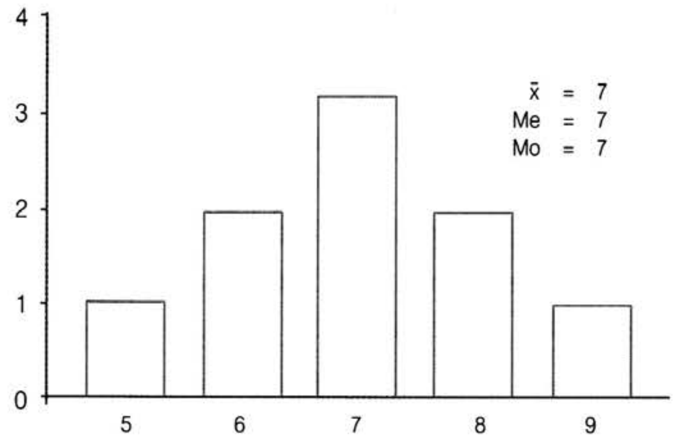


Figura 8.5. Representación gráfica de una distribución simétrica

Suponga ahora que se eliminan los valores 5 y 6 y se agregan el 11 y el 18. ¿Cuál es la nueva distribución y cuáles son la mediana, la moda y la media aritmética?

$$6, 7, 7, 7, 8, 8, 9, 11, 18.$$

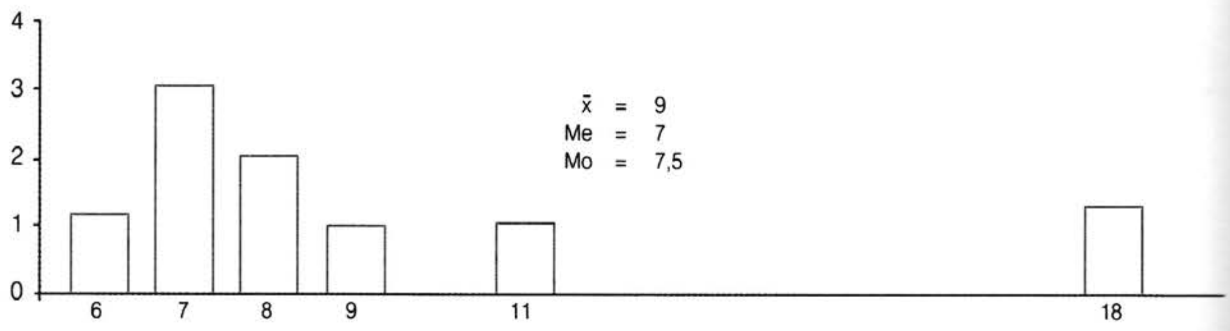


Figura 8.6. Representación gráfica de una distribución asimétrica hacia la derecha

El cambio de dos valores bajos por dos situados a la derecha, uno de ellos muy elevado, hizo la distribución asimétrica hacia la derecha, corriendo un poco la mediana a la derecha, aumentando el promedio significativamente y alejándolo mucho de la moda, la cual no sufrió cambio alguno.

Es evidente que el promedio aritmético, debido a su definición:

$$\frac{\text{Suma de valores}}{\text{Número de valores}}$$

es muy sensible a la presencia de valores extremos. Cuando son altos, el promedio tiende a ser mayor que la mediana y la moda y, cuando son bajos, ocurre el fenómeno inverso.

Esto no sucede con la moda, la cual, por corresponder al valor de la característica que se repite con mayor frecuencia, no es afectada por la presencia de valores extremos, ya sean estos altos o bajos.

A la mediana, que depende del número de datos más que de sus valores, la afecta la asimetría, pero en forma mucho más moderada que al promedio aritmético.

Por lo anterior, cuando las distribuciones son bastante asimétricas, es preferible utilizar la moda y, en especial, la mediana, en lugar del promedio aritmético para caracterizar el grupo de datos. Ahora bien, si en alguna circunstancia se tiene interés en que la medida refleje el efecto de los valores extremos, entonces sí debe emplearse el promedio aritmético, sin importar el grado de asimetría de la distribución.

El mayor uso de la media aritmética en investigaciones, experimentos y, en general, en la vida diaria, se debe entre otras cosas a la popularidad de su concepto y al hecho de presentar más estabilidad que las otras medidas de posición en muestras sucesivas.[6]

Además, en una gran cantidad de situaciones prácticas, interesa estimar totales y la media aritmética es la única adecuada para ese fin porque, al multiplicarla por el número de datos, reproduce la suma de esos datos; esta propiedad se deriva de su definición, como ya fue indicado.

Ejemplo 8

En un hospital ingresaron, en un año, 200 pacientes con el diagnóstico "anemia". El médico director pidió los datos sobre los días de internamiento de esos pacientes en el centro, encontró, al analizarlos, una estancia de 3 días para el que había estado menos y una de 37 para el de mayor permanencia. La suma total de las estancias resultó de 2420 días. Además, observó que la más frecuente era de 8 días y el 50% de los pacientes permaneció en el hospital más de 10 días.

a) ¿Cuál es el recorrido?

b) ¿Cuál es la moda y cuál es la mediana? ¿Por qué?

6. A su vez, la mediana tiene como ventajas sobre la moda: una mayor estabilidad en muestras sucesivas y una menor sensibilidad a los criterios usados para definir los intervalos y elaborar las distribuciones de frecuencias.

c) ¿Cuál es la media aritmética? ¿Cómo lo sabe?

d) Si usted fuera a realizar un estudio del costo de los servicios de ese hospital, ¿qué medida de posición utilizaría? ¿Por qué? Y si fuera usted un médico dedicado a la investigación de las características de la anemia, ¿cuál debería preferir? ¿Por qué?

Solución

a) Recorrido: 37 – 3 = 34 días.

b) $M_o = 8$ días (estancia más frecuente); $M_e = 10$ días (50% de los pacientes permanecieron en el hospital más de 10 días).

c) $\bar{X} = \dfrac{2420}{200} = 12,1$ días (la suma de las estancias es 2420; el número total de pacientes, 200).

d) La media aritmética para el estudio de los costos de los servicios del hospital, ya que interesa considerar los valores extremos. La moda o la mediana para describir la duración típica de la enfermedad y la estancia más común de los pacientes que la sufren.

8.7. LA MEDIA GEOMÉTRICA Y LA MEDIA ARMÓNICA

Otras dos medidas de tendencia central, que se utilizan con cierta frecuencia en algunos campos específicos de la estadística aplicada, son la media geométrica y la armónica. A continuación, se definen y ejemplifica su cálculo.

8.7.1. Media geométrica

Si se tienen n valores de una variable:

$$x_1,\ x_2,\ x_3,\ \ldots,\ x_n.$$

La media geométrica es la raíz n del producto de esos n valores.

Simbólicamente:

$$\bar{X}_g = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n} = (x_1 \cdot x_2 \cdot x_3 \cdot \ldots \cdot x_n)^{\frac{1}{n}}.$$

Ejemplo 9

Obtenga la media geométrica de los números 3, 4, 9 y 12.

$$\bar{X}_g = \sqrt[4]{3 \cdot 4 \cdot 4 \cdot 12} = \sqrt[4]{1296} = \sqrt[4]{6^4} = 6.$$

El cálculo del valor puede hacerse con una calculadora de bolsillo o con la función correspondiente de una hoja de cálculo. En el caso de Excel puede utilizarse la función MEDIA.GEOM. En este caso particular, si se suministra a la hoja de cálculo los valores 3, 4, 9 y 12 y se llama la función MEDI.GEOM, el programa devuelve el valor 6.

Es importante señalar que la media geométrica es la forma correcta de promediar tasas de cambio, relativos de precios, índices y tasas de crecimiento. También es la medida apropiada cuando la distribución es de naturaleza logarítmica, es decir, cuando es acotada a la izquierda y marcadamente asimétrica hacia la derecha, tal como sucede con las distribuciones de ingreso, salarios y aumentos de precios. Por estos motivos tiene mucha aplicación en las áreas de la economía y los negocios.

Presenta, sin embargo, una limitación importante, derivada del hecho de que para calcularla todos los valores deben ser positivos, ya que si uno es 0 o algunos son negativos, su valor es 0 o un número imaginario, entonces carece de todo interés práctico.

Ejemplo 10

Los partidos cristianos tienen décadas de participar electoralmente en Costa Rica. En el cuadro siguiente, se presentan los votos para diputados obtenidos por ellos en las elecciones realizadas desde 1986 hasta febrero del 2010. Los partidos considerados son: Alianza Nacional Cristiana, que participó entre 1986 y el 2002; el partido Renovación Costarricense (de 1998 en adelante) y Restauración Nacional (a partir del 2006); así como Restauración Alajuelense y Restauración Herediana, que lo hicieron en el 2010.

¿En cuánto estimaría el porcentaje de aumento promedio por elección que ha experimentado el voto para diputados por los partidos cristianos entre 1986 y el 2010?

Cuadro 8.8

TOTAL DE VOTOS PARA DIPUTADOS RECIBIDOS POR LOS PARTIDOS CRISTIANOS ENTRE 1986 Y EL AÑO 2010

ELECCIONES DEL AÑO	NÚMERO VOTOS VÁLIDOS	CAMBIO CON RESPECTO A ELECCIÓN ANTERIOR	PORCENTAJE DE CAMBIO
1986	19 972		
1990	22 154	1109	10,9
1994	21 064	0951	-4,9
1998	37 068	1760	76,0
2002	61 524	1660	66,0
2006	88 707	1442	44,2
2010	117 931	1329	32,9

Solución

a) El primer paso es dividir el número de votos válidos recibido en una elección entre los de la elección previa. Así 22 154 entre 19 972 arroja un valor de 1,109, lo cual señala que el voto por los partidos cristianos fue un 110,9% en 1990, comparado con los recibidos en 1986, o sea, que entre 1986 y 1990 el voto cristiano subió en 10,9%. El valor 0,951 obtenido para 1994 indica que la votación recibida en ese año fue un $(1-0{,}951)\cdot 100 = 4{,}9\%$ más baja que la obtenida en 1990.

b) Luego, se procede a calcular la media geométrica de los cambios aplicando la fórmula correspondiente:

$$MG = \sqrt[6]{(1{,}109)(0{,}951)(1{,}760)(1{,}660)(1{,}442)(1{,}329)} = 1{,}3444.$$

Si se le resta 1 y se multiplica por 100, se obtiene una tasa de crecimiento promedio entre elecciones de 34,44% para el voto cristiano para diputados. Este es el porcentaje correcto de aumento promedio del voto entre períodos.

c) Es importante notar que la media aritmética de los cambios es:

$$\bar{X} = \frac{10{,}9 - 4{,}9 + 76 + 66 + 44{,}2 + 32{,}9}{6} = \frac{225{,}1}{6} = 37{,}52.$$

Esto señala un porcentaje de cambio de 37,52%, valor superior al arrojado por la media geométrica.

Ahora bien, si se toma el valor inicial de la serie: 19 972 votos, y se le aplica sucesivamente la tasa de crecimiento 37,52, arrojada por la media aritmética, hasta llegar al 2010, se obtiene una estimación de 135 088 votos:

$$19\,972\,(1{,}3752)\,(1{,}3752)\,(1{,}3752)\,(1{,}3752)(1{,}3752)\,1{,}3752) = 135\,088$$

valor significativamente superior a los 117 931 recibidos en la elección.

En cambio, si se realiza el mismo ejercicio usando la tasa promedio arrojada por la media geométrica, el valor resultante para el 2010 es:

$$19\,972\,(1{,}3444)\,(1{,}3444)\,(1{,}3444)\,(1{,}3444)\,(1{,}3444)\,(1{,}3444) = 117\,921$$

el cual es muy cercano a los 117 931 recibidos en la elección del año 2010.

8.7.2. Media armónica

Si se tiene un conjunto de n valores: $x_1, x_2, x_3, \ldots, x_n$.

La media armónica se define como el recíproco de la media aritmética de los recíprocos de los valores. En símbolos: $x_a = \dfrac{1}{\dfrac{\frac{1}{x_1}+\frac{1}{x_2}+\frac{1}{x_3}+\ldots+\frac{1}{x_n}}{n}} = \dfrac{1}{\dfrac{\sum_{i=1}^{n}\frac{1}{x_i}}{n}}$

o sea, $x_a = \dfrac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}$.

Ejemplo 11

Calcular la media armónica de los valores: 3,2, 6.

$$\overline{x_a} = \frac{3}{\frac{1}{3}+\frac{1}{2}+\frac{1}{6}} = \frac{3}{\frac{2+3+1}{6}} = \frac{3\cdot 6}{6} = 3.$$

La media armónica de los números resulta ser 3.[7]

◇◇◇

Ejemplo 12

Como ilustración adicional, considere un caso en el que se desea promediar velocidades.

Un hombre hace un recorrido de 600 km por un camino donde se dan variaciones importantes en su estado de mantenimiento. Él viaja 200 km a 40 km por hora, otros 200 km a 50 km por hora y, finalmente, 200 km más a una velocidad de 100 km por hora. ¿Cuál es la velocidad promedio para los 600 km recorridos?

Cuando se nota que todas las distancias recorridas a diferentes velocidades son las mismas, 200 km, parece que la solución correcta se obtiene calculando el promedio aritmético de las tres velocidades, o sea:

$$\frac{40+50+100}{3} = 63{,}33\ \text{Km/h}.$$

Sin embargo, este resultado no es correcto, como puede apreciarse seguidamente:

$$\text{Velocidad} = \frac{\text{distancia}}{\text{tiempo}};\quad v = \frac{d}{t} \text{ y } t = \frac{d}{v}.$$

7. El cálculo puede hacerse directamente utilizando la función MEDIA.ARMO de la hoja de cálculo Excel.

La distancia total recorrida es 600 km y el tiempo consumido en cada uno de los tramos del recorrido se obtiene con la división de la distancia entre la velocidad para cada uno de los tres casos y luego se suman los resultados:

$$t_1 = \frac{200}{40} = 5 \text{ horas}, \; t_2 = \frac{200}{50} = 4 \text{ horas}, \; t_3 = \frac{200}{100} = 2 \text{ horas}.$$

$$\text{tiempo total} = t_1 + t_2 + t_3 = 5 + 4 + 2 = 11 \text{ horas}.$$

$$\text{velocidad} = \frac{600}{11} = 54,5 \text{ k/h}.$$

La velocidad media para todo el recorrido es, entonces, de 54,5 km por hora. A este mismo resultado se puede llegar calculando la media armónica de las velocidades.[8]

$$x_a = \frac{3}{\frac{1}{40} + \frac{1}{50} + \frac{1}{100}} = \frac{3}{\frac{5+4+2}{200}} = \frac{3 \cdot 200}{11} = \frac{600}{11} = 54,5.$$

◇◇

8.7.3. Relación entre la media aritmética, la geométrica y la armónica

Resulta interesante establecer una relación de orden de magnitud entre las tres medias estudiadas.

Dada una serie de números: $x_1, x_2, x_3, \ldots, x_n$, la media geométrica es menor o igual que la media aritmética pero mayor o igual a su media armónica. En símbolos, esta relación es la siguiente:

$$\bar{X}_a \leq \bar{X}_g \leq \bar{X}.$$

Las tres medidas tendrán igual valor solamente cuando todos los números de la serie sean idénticos.

8. O también calculando la media ponderada (por los tiempos) de las velocidades: $40\,5 + 50 \cdot 4 + 200 \cdot 29/11 = 54{,}5$.

EJERCICIOS DE AUTOEVALUACIÓN

1. Indique cuál es el propósito que deben cumplir las medidas de tendencia central y qué aspectos del conjunto de datos ayudan a conocer principalmente.

2. ¿Cuáles son las tres medidas de tendencia central de mayor importancia?

3. Usando las propiedades del símbolo de sumatoria, calcule el valor numérico de las expresiones que seguidamente se indican:

$$\sum x,\ \sum y,\ \sum x^2,\ \sum xy,\ \left(\sum x\right)^2 \cdot \sum y,\ \sum (x-y),\ \sum (x-y)\cdot \sum (x-y);$$

$$x_1 = 1;\quad x_2 = -3;\quad x_3 = -2;\quad x_4 = 3;\quad x_5 = 3$$

$$y_1 = 2;\quad y_2 = 2;\quad y_3 = 1;\quad y_4 = 0;\quad y_5 = 1$$

4. La sumatoria de la diferencia de dos variables $\sum (x-z)$ es igual a

5. La medida de tendencia central que toma en cuenta, en mejor forma, el valor típico o que más se repite en un conjunto de datos se denomina

6. Si de un conjunto grande de datos se escoge uno al azar, ¿a qué valor tiene más probabilidad de ser igual el valor observado: al de la media aritmética, al de la moda o el de la mediana?

7. Si en un examen de contabilidad la parte teórica tiene el doble de puntaje que la parte práctica, qué nota le corresponde a un estudiante que obtiene un 8 en teoría y un 5 en la parte práctica.

8. ¿Cuáles son las medidas de posición que coinciden (valen lo mismo) en toda distribución o curva simétrica?

9. Cuando las distribuciones de frecuencias tienen clases desiguales, ¿cuál de las siguientes fórmulas puede aplicarse?

   a) La de **la moda**, como si se tratara de clases de igual amplitud.

   b) La de **la mediana**, ajustando las frecuencias absolutas y las acumuladas.

   c) La de **la media aritmética**, sin cambio alguno.

10. Usando los valores 1, 4, 5, 9, 11 calcule la media aritmética y luego compruebe empíricamente las cuatro propiedades enunciadas del promedio aritmético. Utilice $b = 1{,}25$.

11. Los datos que se presentan a continuación corresponden a notas de aprovechamiento de un grupo de 30 estudiantes de un curso de verano de Administración, en la Universidad de Costa Rica:

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

6,9 7,0 7, 0 7, 0 7, 0 7,6 7, 6 7,9 7, 9 8,2 8,4 8,7 8, 7

8,7 9,2 9,5 9, 5

a) Calcule la media aritmética utilizando la definición de promedio simple.

b) Calcule la media aritmética utilizando las notas diferentes y la definición de promedio ponderado.

12. En un estudio en el que se investigó el salario por hora de obreros calificados, se obtuvieron los siguientes valores (en colones): promedio 1350, mediana 1100 y moda 1030.

a) Haga un bosquejo de la posible distribución de los salarios por hora de los obreros especializados, a partir de la información disponible sobre las principales medidas de posición. Preste especial atención al punto de la asimetría. No olvide marcar, donde corresponda en el bosquejo, la mediana, la moda y la media aritmética.

b) ¿Cuál de esas medidas utilizaría para representar los salarios de los trabajadores?

13. Hallar la media geométrica y la media armónica de los siguientes números:

5, 12, 17, 28, 33

14. Un hombre viaja en su auto de la ciudad de Alajuela a la de Heredia a una velocidad media de 30 km/h y vuelve de Heredia a Alajuela por la misma ruta, con una velocidad media de 60 km/h. Hallar la velocidad media para el viaje completo.

15. Un arquitecto ha sido contratado para diseñar las viviendas que servirán de modelo para una gran urbanización. Uno de los problemas que debe resolver es el tamaño de las cocheras o garajes, más concretamente si debe diseñarlas para un carro, para dos o más grandes. Al indagar sobre el tema, descubre un estudio de mercadeo relativamente reciente donde se observó una población de nivel socioeconómico similar y se obtuvo información sobre el número de vehículos por familia. Los datos señalan: $x = 2{,}15$, Med = 1,4 y Mo = 2.

a) ¿En cuál de las medidas se basará para decidir el tamaño de las cocheras de las casas tipo?

b) ¿Por qué preferiría esa medida?

16. Suponga que en una institución académica se abre una plaza que se llenará por concurso. Cada aspirante será evaluado en cuatro áreas: años de experiencia académica (AEA), nivel de estudios (Bachillerato, Maestría, Doctorado) (NES), publicaciones realizadas (PUB) y personalidad (una entrevista, ENT). En cada una de esas áreas recibirá una nota entre 1 y 10.

    El puntaje final se obtendrá dando los siguientes pesos a las notas: AEA: 20%, NES: 10%, PUB: 30% y ENT: 40%.

    El aspirante Hugo Rafael Morales obtiene las siguientes calificaciones:

    AEA: 7, NES: 6, PUB: 4 y ENT: 8. ¿Cuál es el puntaje final de Hugo Rafael?

RESPUESTA A LOS EJERCICIOS DE AUTOEVALUACIÓN

1. El propósito que deben cumplir las medidas de posición es el de resumir, en un solo número, la posición o localización de la distribución; y el aspecto del conjunto de datos que nos ayudan a conocer es la posición de la distribución, o sea, alrededor de qué valor se tienden a concentrar los datos; por ello, a las medidas de posición se les llama también valores centrales.
2. Las tres medidas de posición de mayor importancia son: la media aritmética o promedio, la mediana y la moda.
3. $\sum x = 1 + -3 + -2 + 3 + 3 = 2$

   $\sum y = 2 + 2 + 1 + 0 = 6$

   $\sum x^2 = 1^2 + (-3)^2 + (-2)^2 + 3^2 + 3^2 = 1 + 9 + 4 + 9 + 9 = 32$

   $\sum xy = 1 \cdot 2 + -3 \cdot 2 + -2 \cdot 1 + 3 \cdot 0 + 3 \cdot 1 = 2 - 6 - 2 + 0 + 3 = -3$

   $(\sum x)^2 \sum y = 2^2 \cdot 6 = 24$

   $\sum (x - y) = (1 - 2) + (-3 - 2) + (-2 - 1) + (3 - 0) + (3 - 1) = -4$

   $\sum (x - y) \sum (x - y) = -4 \cdot -4 = 16$
4. La sumatoria de la diferencia de dos variables es igual a la suma algebraica de la sumatoria de las variables, así: $\sum (x - z) = \sum x - \sum z$.
5. Moda. Se simboliza Mo.
6. Al valor de la moda, pues corresponde al valor que más se repite y siempre es un valor observado. La probabilidad de que se parezca a la mediana es bajo, ya que esta es solo un valor central y, en muchos casos, su valor se obtiene promediando los dos valores centrales observados. Sería, finalmente, una gran casualidad, por no decir imposible, que se parezca a la media, porque por su forma de cálculo casi nunca coincide con un valor observado.

7. La nota promedio es de 7 porque si la parte teórica vale el doble, se debe ponderar por 2 y la parte práctica, por 1; así:

$$\frac{8 \cdot 2 + 5 \cdot 1}{2 + 1} = \frac{16 + 5}{3} = 7.$$

8. Las medidas de posición que coinciden (valen lo mismo) en toda distribución o curva simétrica son: la media aritmética, la mediana y la moda.

9. Opción c).

10.

$$\overline{X} = \frac{\sum x_i}{n} = \frac{1 + 4 + 5 + 9 + 11}{5} = \frac{30}{5} = 6.$$

Propiedad (1)

$\sum x_i = nx = 5 \cdot 6 = 30 = 1 + 4 + 5 + 9 + 11.$

Propiedad (2)

$$\begin{aligned}\sum (x_i - x) &= (1 - 6) + (4 - 6) + (5 - 6) + (9 - 6) + (11 - 6) \\ &= -5 + -2 + -1 + 3 + 5 = -8 + 8 = 0.\end{aligned}$$

Propiedad (3)

Si $y_i = x_i \pm b$, entonces $\overline{y} = \overline{x} \pm b$.

Si se resta $b = 1,25$, entonces:

$$\begin{aligned}\overline{y} &= \frac{(1 - 1,25) + (4 - 1,25) + (5 - 1,25) + (9 - 1,25) + (11 - 1,25)}{5} \\ &= \frac{23,75}{5} = 4,75 = 6 - 1,25.\end{aligned}$$

Propiedad (4)

Si $y_i = x_i b$, entonces $\overline{y} = \overline{x} b$.

Si se multiplica por $b = 1,25$, entonces:

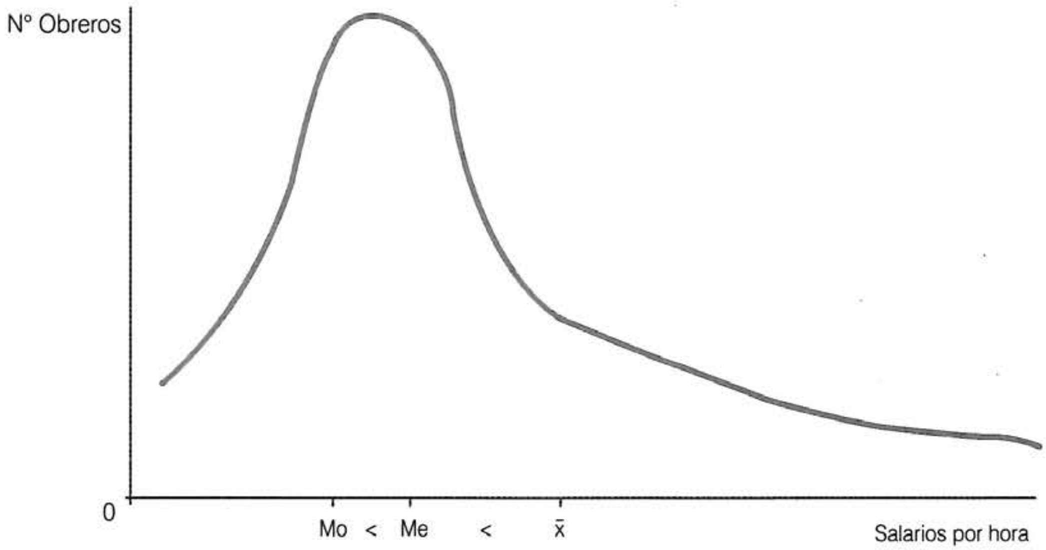
$$\begin{aligned}y &= \frac{1 \cdot 1,25 + 4 \cdot 1,25 + 5 \cdot 1,25 + 9 \cdot 1,25 + 11 \cdot 1,25}{5} \\ &= \frac{37,5}{5} = 7,5 = 6 \cdot 1,25.\end{aligned}$$

11. a) $\dfrac{\sum x_i}{n} = \dfrac{1,0 + 2,3 + 4,0 + ... + 9,5}{30} = \dfrac{200}{30} = 6,67.$

b)

1,0 · 1 = 0	6,1 · 1 = 6,1	8,2 · 1 = 8,2
3,1 · 1 = 2,3	6,3 · 1 = 6,3	8,4 · 1 = 8,4
3,1 · 1 = 3,1	6,8 · 3 = 20,4	8,7 · 3 = 26,1
4,0 · 2 = 8,0	6,9 · 1 = 6,9	9,2 · 1 = 9,2
4,7 · 1 = 4,7	7,0 · 4 = 28,0	9,5 · 2 = 19,0
5,3 · 1 = 5,3	7,6 · 2 = 15,2	Total 200,0
6,0 · 1 = 6,0	7,9 · 2 = 15,8	

12. a)



b) Para representar lo típico, lo más adecuado será utilizar la moda, aunque también, en este caso, podría ser la mediana. Si se requiere calcular el total de salarios pagados o la cantidad recibida por un grupo de obreros, entonces sería útil el promedio.

13. Se utiliza la fórmula $\overline{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot ... \cdot x_n}$ para la media geométrica:

$$\overline{x}_g = \sqrt[5]{5 \cdot 12 \cdot 17 \cdot 28 \cdot 33} = \sqrt[5]{942480} = 15,662.$$

Para la media armónica se usa:

$$\overline{x}_a = \frac{1}{\dfrac{\dfrac{1}{x_1} + \dfrac{1}{x_2} + ... + \dfrac{1}{x_n}}{n}}$$

$$\overline{x}_a = \frac{1}{\dfrac{\dfrac{1}{5} + \dfrac{1}{12} + \dfrac{1}{17} + \dfrac{1}{28} + \dfrac{1}{33}}{5}} = \frac{1}{\dfrac{0,408}{5}} = \frac{1}{0,0816} = 12,255.$$

$$u_i = \frac{x_i - x_0}{c}.$$

PUNTOS MEDIOS	FRECUENCIAS	PUNTOS MEDIOS CORREGIDOS	
462	98	-4	-392
480	75	-3	-225
498	56	-2	-112
516	42	-1	-42
534	30	0	0
552	21	1	21
570	15	2	30
588	11	3	33
606	6	4	24
624	2	5	10
Total	356		-653

Cálculo de $\overline{u}$ se utiliza la fórmula corriente del promedio

$$\overline{u} = \frac{\sum u_i f_i}{n} = \frac{-635}{356} = -1,8343.$$

Y el cálculo de $\overline{x}$ se realiza mediante $\overline{x} = c\overline{u} + x_0$

$$\overline{x} = 18(-1,8343) + 534 = 500,98.$$

14. Suponga que la distancia de la ciudad de Alajuela a la ciudad de Heredia es de 12 km (aunque puede ser supuesta cualquier distancia). Entonces:

Tiempo para ir de Alajuela a Heredia $= \frac{d}{v} = \frac{12 \text{ km}}{30 \text{ km/h}} = 0,4 \text{ h}.$

Tiempo para ir de Heredia a Alajuela $= \frac{d}{v} = \frac{12 \text{ km}}{60 \text{ km/h}} = 0,2 \text{ h}.$

Velocidad media para todo el viaje $= \frac{d \text{ total}}{t \text{ total}} = \frac{24 \text{ km}}{0,6 \text{ h}} = 40 \text{ km/h}.$

El promedio anterior es la media armónica de 30 y 60, es decir,

$$\overline{x}_a = \frac{1}{\frac{\frac{1}{30} + \frac{1}{60}}{2}} = \frac{1}{\frac{0,05}{2}} = \frac{1}{0,025} = 40 \text{ km/h}.$$

15. a) Se usa la moda.

    b) Es bastante obvio que la media y la mediana no representan realmente el número de carros poseídos por las familias. La moda representa el tamaño de garaje que llenaría las necesidades del mayor número de familias.

16. El puntaje final de Hugo Rafael debe calcularse haciendo una media ponderada de las calificaciones logradas en los diferentes rubros y usando como ponderaciones los pesos asignados a ellas:

ÁREA	NOTA RECIBIDA (x_i)	PESO ASIGNADO (w_i)	$x_i \cdot w_i$
Experiencia académica	7	0,20	1,40
Nivel estudios	6	0,10	0,60
Publicaciones realizadas	4	0,30	1,20
Personalidad (Entrevista)	8	0,40	3,20
TOTAL		1,00	6,40

$$\text{Puntaje final} = \frac{\sum x_i \cdot w_i}{\sum w_i} = \frac{6,40}{1} = 6,40.$$