ESTADÍSTICA
GUIA No 4

MEDIDAS DE CENTRALIZACIÓN Y LOCALIZACIÓN

Introducción

La elaboración de tablas de distribución de frecuencias hace parte, como se dijo en el capítulo anterior, del proceso de reducción de los datos estadísticos. La máxima expresión del resumen de todos los datos observados, relativos a una variable, se obtiene al elegir un único valor representativo o típico de la distribución.

Cuando los datos observados corresponden todos a una población de estudio, este valor representativo o típico, que es fijo para cada población, recibe el nombre de **parámetro**. Sin embargo, lo más usual es trabajar con datos de una o varias muestras dicho valor representativo, que al referirse a una muestra recibe el nombre de **estadístico** o **estadígrafo**. Lo habitual es que el valor de un estadístico varíe de una a otra muestra acercándose más menos al valor del parámetro. Justamente es la estadística inferencial o inductiva la que establece las condiciones bajo las cuales el estadístico representa al parámetro que pretende estimar.

En realidad, los resúmenes numéricos o estadísticos, surgen de la necesidad de modelar la variación estadística en diferentes sentidos: localización, dispersión, forma (asimetría y curtosis) y concentración. Para ello se suele considerar diferentes tipos de estadísticas o estadígrafos que concretan cada una de estos conceptos en términos de una medición estadística.

En este capítulo centraremos la discusión en las medidas de localización y trataremos de responder a cuestiones como qué son, cuáles son, cómo se definen, qué propiedades, ventajas e inconvenientes las caracterizan y en qué tipos de problemas o situaciones se aplican.

Las **medidas de localización**, también conocidas como medidas de posición, son un conjunto de estadígrafos con los que se busca identificar un valor que resuma, represente o caracterice una posición o tendencia particular de un conjunto de datos. Algunos ejemplos de medidas de localización son: el máximo de los datos, el mínimo de los datos, la media aritmética, los cuantiles y la moda. El caso particular y más relevante de las medidas de localización lo constituyen las **medidas de tendencia central** que buscan identificar un valor intermedio, más característico o de posición central de los datos. De los ejemplos nombrados antes sólo la media aritmética y la moda son medidas de tendencia central. Los procedimientos de cálculo y las características de estas medidas varían de acuerdo al tipo de variables que se quiera resumir y a la cantidad de información que se trabaje.

En lo que sigue, primero se hará una descripción de las medidas de tendencia central más conocidas, luego se comentarán otras medidas de localización como los cuartiles y finalmente se presentará una serie de ejemplos, para ilustrar el cálculo y la elección apropiada de los diferentes estadísticos de localización.

4.1 Media aritmética.

La media aritmética denotada por $\overline{X}$, es quizás la medida de tendencia central más utilizada, conocida y sencilla de calcular. Además, es una medida de gran estabilidad en el muestreo y sus fórmulas admiten tratamientos algebraicos. Su principal desventaja es la de ser muy sensible a cambios que se hagan en alguno de sus valores, como por ejemplo cuando los valores extremos son valores demasiado grandes o pequeños.

La media aritmética es el número obtenido al dividir la suma de todos los valores de la variable entre el número total de observaciones, y se puede definir de manera general, a través de la expresión $\overline{X} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i}$.

En realidad esta expresión usualmente se utiliza cuando se trabaja con datos agrupados en donde los n_i representan las frecuencias absolutas y los x_i representan los diferentes valores que asumen los datos, o bien corresponden a las marcas de clase m_i de tablas de variable continua. Entonces, el número de marcas de clase o de valores diferentes es k y el total de datos es $\sum_{i=1}^{k} n_i$. Cuando no se trabaja con datos agrupados se entiende que $n_i = 1$ para toda i, que $\sum_{i=1}^{k} n_i = n$ el total de datos y la media aritmética simplemente se denota como $\overline{X} = \frac{\sum_{i=1}^{n} x_i}{n}$.

En cuanto a las propiedades que caracterizan a la media aritmética tenemos:

- Si los datos x_i se transforman en datos de la forma $y_i = a + b\, x_i$, entonces se tiene que $\overline{Y} = a + b\,\overline{X}$.
- La suma de las desviaciones de los valores de una variable x, respecto a su media aritmética es cero. Es decir, $\sum_{i=1}^{n} (x_i - \overline{x}) = 0$.
- La suma de las desviaciones al cuadrado es mínima en $\overline{X}$, es decir, $\overline{X}$ es el valor que minimiza la función $F(u) = \sum_{i=1}^{n} (x_i - u)^2$.

Respecto a sus ventajas se pueden citar las siguientes:

- Es quizás, la medida más fácil de entender, la más usada y viene expresada en las mismas unidades que la variable.
- Es un número comprendido entre el menor y el mayor de los valores a los que se aplica.
- En su cálculo intervienen todos los valores de la distribución de datos.
- Representa, en términos físicos, el centro de gravedad de toda la distribución de datos.

- Es única, ya que $\overline{X}$ sin agrupar se puede ver como una función de R^n en R y de manera similar para $\overline{X}$ agrupada.
- Es muy estable en el muestreo de datos.
- Es altamente sensible a cualquier cambio en los datos de la distribución.
- Es adaptable cuando se trata de hacer cálculos matemáticos posteriores con ésta, como en el caso del promedio ponderado o del promedio de promedios.

En cuanto a sus inconvenientes se tiene que:

- Se ve afectada por los valores extremadamente grandes o pequeños de la distribución de los datos. Por ello, la inclusión de valores atípicos en la distribución que se esté estudiando, puede dar una media aritmética que no sea realmente un representante típico del grupo.
- Cuando una distribución de datos es marcadamente asimétrica, casos donde la media aritmética, la mediana y la moda difieren en forma apreciable, debe considerarse la posibilidad de que pueda no ser el único valor representativo de los datos.
- Cuando la distribución de datos tiene forma de U, la media aritmética corresponde a los valores menos comunes de los datos y por tanto, puede dar una idea irreal de la distribución.

4.2 Media aritmética ponderada (W).

La media aritmética ponderada, denotada como W, es una variación de la media aritmética que se emplea en distribuciones de tipo univariado, en las que se introducen unos coeficientes de ponderación, denominados w_i, que son valores positivos que representan el número de veces que un valor de la variable es más importante que otro.

$$W = \frac{\sum_{i=1}^{k} x_i w_i}{\sum_{i=1}^{k} w_i}$$

En general la media aritmética ponderada goza de las mismas propiedades, ventajas e inconvenientes de la media aritmética, ya que se puede equiparar con una media aritmética agrupada.

4.3 Media geométrica (G).

Para una distribución de frecuencias definida en términos de k parejas (x_i, n_i), la media geométrica, que denotaremos como G, se define como la raíz n-ésima del producto de los n valores, es decir $G = \sqrt[n]{x_1^{n_1} x_2^{n_2} x_3^{n_3} \ldots x_k^{n_k}}$.

Si se utilizan datos sin agrupar se reduce simplemente a $G = \sqrt[n]{x_1 x_2 x_3 \ldots x_n}$.

El empleo más frecuente de la media geométrica es el de promediar mediciones tales como porcentajes, tasas, números índices. etc., es decir, en los casos en los que se supone que la variable presenta variaciones acumulativas.

En cuanto a las ventajas e inconvenientes de la media geométrica se destaca lo siguiente:

- En su cálculo intervienen todos los valores de la distribución.
- Su valor está comprendido entre el menor y el mayor de los datos.
- Ante la presencia de valores extremos tienen menor influencia que en el caso de la media aritmética.
- Es única, ya que se puede ver como una función de R^n en R
- Su cálculo es más complicado que el de la media aritmética.
- No cambia cuando se reordenan los valores de la variable y cumple la propiedad de homogeneidad.
- Cuando la variable toma al menos un valor $x_i = 0$ entonces G se anula, y si la variable toma valores negativos se pueden presentar una gama de casos particulares en los que tampoco queda determinada debido al problema de las raíces de índice par para números negativos.

4.4 Media armónica (H).

Hay ocasiones en que los valores de una variable vienen expresados en términos de otra que es inversamente proporcional o recíproca de la primera como en el caso de la velocidad y el tiempo o de la demanda de un bien y su precio de mercado. En estos casos se necesita un promedio que tenga en cuenta la reciprocidad. La media armónica, que se denotará como H, satisface estos requerimientos y se define así:

$$H = \frac{N}{\sum_{i=1}^{k} \frac{1}{x_i} n_i}$$

Para datos sin agrupar la media armónica simplemente será:

$$H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

En cuanto a las ventajas e inconvenientes de la media armónica se mencionan las siguientes:

- Es única y en su cálculo intervienen todos los valores de la distribución.
- Su valor siempre estará comprendida entre el menor y el mayor de los valores de la variable.
- Su cálculo no tiene sentido cuando algún valor de la variable toma el valor cero.
- Su uso no es recomendable en distribuciones de variables con valores pequeños.

4.5 Error medio cuadrático (Emc).

Esta medida, que se denota como Emc, se usa raramente como medida de tendencia central. Por ejemplo, se utiliza cuando se quiere dar un estimativo del error cometido en una medición en donde hay valores negativos y positivos y se quiere prescindir de si estos errores fueron por exceso o por defecto.

Para datos agrupados se define como $\text{Emc} = \sqrt{\frac{1}{n}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}$.

Para datos sin agrupar se define como $\text{Emc} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$.

En cuanto a ventajas e inconvenientes cabe señalar los siguientes:

- Es demasiado sensible a la oscilación de valores extremos.
- Es invariante ante una reordenación de los datos.
- Satisface la propiedad de homogeneidad

En el ejemplo 14 de la próxima sección, se comentan más detalles acerca de esta medida, y de la forma como se define.

4.6 Media potencial (Mp).

Se llama **media potencial** de grado **p** de los números $x_1, x_2, \ldots, x_n$ al número Mp definido como $\text{Mp} = \left(\frac{x_1^{\,p} + x_2^{\,p} + \ldots + x_n^{\,p}}{n}\right)^{1/p}$.

En particular:

Si p=2, el número $M_2 = \left(\frac{x_1^{\,2} + x_2^{\,2} + \ldots + x_n^{\,2}}{n}\right)^{1/2}$ se llama **media cuadrática**.

Si p=-1, el número $M_{-1} = \left(\frac{x_1^{\,-1} + x_2^{\,-1} + \ldots + x_n^{\,-1}}{n}\right)^{-1} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \ldots + \frac{1}{x_n}}$ se llama **media armónica**.

4.7 Mediana (Me).

Dada una distribución de frecuencias con los valores ordenados de menor a mayor, se llama mediana y se representa por Me, al valor de la variable, que deja a su izquierda el mismo número de frecuencias que a su derecha, o en términos más sencillos, la mediana es el valor que divide al conjunto en dos partes iguales, de tal forma que el número de valores mayor o igual a la mediana es igual al número de valores menores o igual a ésta. Para el cálculo de la mediana se contemplan tres casos.

En primer lugar, la determinación del valor de la mediana para variables discretas no agrupadas de un conjunto de n datos se puede realizar así: primero se ordenan los datos en una tabla de frecuencias y se calcula n/2, luego se construye la columna de las frecuencias acumuladas (N_i) y entonces se observa cual es la primera N_i que supera o iguala a n/2 distinguiéndose dos casos:

- Si existe un valor x_i tal que $N_{i-1} < n/2 < N_i$, la mediana es $\text{Me} = x_i$.
- Si existe un valor x_i tal que $N_i = n/2$, la mediana es $\text{Me} = (x_i + x_{i+1})/2$.

En segundo lugar, cuando no se tiene a la mano una distribución de frecuencias o no se desea elaborarla, debido por ejemplo, a que hay muy pocos datos, el cálculo de la mediana se puede realizar así: se ordenan los datos de menor a mayor o viceversa y dependiendo del número de observaciones se darán dos casos:

- Si el número de datos es impar, la mediana es igual al dato central, es decir al dato número (n+1)/2.
- Si el número de datos es par, la mediana es igual al promedio aritmético de los datos centrales o a cualquier valor comprendido entre ellos.

Finalmente, cuando se tiene una variable continua agrupada en una tabla de frecuencias, la mediana se puede calcular teniendo en cuenta dos casos:

- Caso 1: cuando $N_{j-1} = \frac{n}{2}$ se tiene $Me = y_{j-1}$
- Caso 2: cuando $N_{j-1} < \frac{n}{2}$ se tiene $Me = y_{j-1} + C\left[\frac{\frac{n}{2} - N_{j-1}}{n_j}\right]$

Donde n es el número de observaciones, C es la amplitud de la clase, y_{j-1} es el extremo inferior de la clase a la cual pertenece la mediana, n_j es la frecuencia absoluta de la clase a la que pertenece la mediana y N_{j-1} es la frecuencia acumulada correspondiente a la clase anterior a la que contiene la mediana.

Una propiedad interesante de la mediana es que la suma de los valores absolutos de las desviaciones tiene su valor mínimo, cuando se consideran las desviaciones con respecto a la mediana, es decir, el valor que minimiza la función G es la mediana donde

$$G(u) = \sum_{i=1}^{n} |x_i - u|$$

En cuanto a las ventajas e inconvenientes se mencionan los siguientes:

- Es la medida más representativa en el caso de variables que sólo admitan la escala ordinal.
- En su cálculo sólo influyen los valores centrales y es insensible a los valores extremos o atípicos, lo cual también se puede ver como una desventaja si no hay valores atípicos.
- En su determinación no intervienen todos los valores de la variable, por lo que no aprovecha toda la información de los datos.
- La mediana no se adapta a cálculos posteriores aritméticos, en la medida en que si obtenemos las medianas de diferentes grupos, no podemos obtener una mediana de los grupos reunidos.

4.8 Cuantiles.

Los cuantiles se pueden ver como una familia general de estadísticos de localización, ya que son aquellos valores de la variable, que ordenados de menor a mayor, dividen a la distribución de datos en partes, de tal manera que cada una de ellas contiene el mismo número de frecuencias. Los cuantiles más conocidos son los Cuartiles (Qi), los Deciles (Di) y los Percentiles (Pi). En lo que sigue se dedicará más atención a los cuartiles, pues, como se reporta en el capítulo anterior, es con base en ellos que se construye e interpreta el Gráfico de caja.

4.8.1 Cuartiles (Q_i).

Con base en estos estadísticos se identifican valores de la variable que dividen la distribución de datos, previamente ordenada de menor a mayor, en cuatro partes, cada una de las cuales engloba el 25% de las mismas. Se denotan de la siguiente forma: Q_1 es el primer cuartil que deja a su izquierda el 25 % de los datos; Q_2 es el segundo cuartil que deja a su izquierda el 50% de los datos (este cuartil es la misma mediana), y Q_3 es el tercer cuartil que deja a su izquierda el 75% de los datos.

En cuanto al cálculo de los cuartiles es curioso y confuso, ver que hay diferentes criterios para determinar los cuartiles. Behar y Grima (2004) comentan e ilustran cuatro métodos de los que se dará cuenta en este apartado: el de Tukey, el de Moore y McCabe, el de Minitab y el de Excel. Se comienza comentado el método de Tukey.

Método de Tukey (Para un número impar de datos)				
2	4	6	8	10
	$Q_1 = 4$	Me = 6	$Q_3 = 8$	
Método de Tukey (Para un número par de datos)				
2	4	--	6	8
$Q_1 = 3$		Me = 5	$Q_3 = 7$	

Tabla 4.1. Ejemplo del método de Tukey para calcular los cuartiles

En la Tabla 4.1 se distinguen dos casos. Cuando hay un número impar de datos (en el ejemplo hay cinco datos) los cuartiles Q_1 y Q_3 coinciden con los valores de los datos ubicados en la posición segunda y cuarta respectivamente, mientras que cuando hay un número par de datos, Tukey propone calcular el promedio de los datos primero y segundo para Q_1 y de los datos tercero y cuarto para Q_3.

Método de Moore y McCabe (Para un número impar de datos)				
2	4	6	8	10
$Q_1 = 3$		Me = 6	$Q_3 = 9$	
Método de Moore y McCabe (Para un número par de datos es igual a Tukey)				
2	4	--	6	8
$Q_1 = 3$		Me = 5	$Q_3 = 7$	

Tabla 4.2. Ejemplo del método de Moore y McCabe para calcular los cuartiles

Cuando hay un número impar de datos, Moore y McCabe prefieren, como se puede ver en la Tabla 4.2, calcular el promedio de los datos primero y segundo para Q_1, y de los datos cuarto y quinto para Q_3. Para el caso de un número par de datos, el método de Moore y McCabe coincide con el de Tukey.

<table>
<tr><td colspan="5">Método de Minitab - Usa “posicionadores” 0,25(n+1) y 0,75(n+1)
Para número par de datos usa como interpoladores
$Q_1 = x_1 + 0,25(x_2-x_1)$ y $Q_3 = x_3 + 0,75(x_4-x_3)$
Para número impar de datos usa como interpoladores
$Q_1 = x_1 + 0,5(x_2-x_1)$ y $Q_3 = x_3 + 0,5(x_5-x_4)$</td></tr>
<tr><td>2</td><td>4</td><td>--</td><td>6</td><td>8</td></tr>
<tr><td colspan="2">$Q_1 = 2,5$</td><td>Me = 5</td><td colspan="2">$Q_3 = 7,5$</td></tr>
<tr><td>2</td><td>4</td><td>6</td><td>8</td><td>10</td></tr>
<tr><td colspan="2">$Q_1 = 3$</td><td>Me = 6</td><td colspan="2">$Q_3 = 9$</td></tr>
</table>

Tabla 4.3. Ejemplo del método de Minitab para calcular los cuartiles

En cuanto a los programas para computadora, como Minitab y Excel, los algoritmos que se emplean utilizan la idea de posicionadores. En particular, Minitab utiliza las expresiones 0,25(n+1) y 0,75(n+1) para identificar las posiciones de Q_1 y Q_3 respectivamente. En la Tabla 4.3, cuando n = 4 se tiene que 0,25 (n+1) = 1,25, lo que indica que el valor de Q_1 estará entre los dato x_1 y x_2, de manera que $Q_1 = x_1 + 0,75(x_2-x_1) =$ 2,5. En cambio en Excel, como se ve en la Tabla 4.4, cuando n = 4 el posicionador da 0,25(n-1) + 1 = 1,75, lo que indica también que el valor de Q_1 estará entre los dato x_1 y x_2, pero en este caso $Q_1 = x_1 + 0,75(x_2-x_1)$ = 3,5.

<table>
<tr><td colspan="5">Método de Excel - Posicionadores: 0,25(n-1) +1 y 0,75(n-1) +1
Para número par de datos usa como interpoladores
$Q_1 = x_1 + 0,75(x_2-x_1)$ y $Q_3 = x_3 +0,25(x_4-x_3)$
Para número impar de datos el posicionador da una posición exacta.</td></tr>
<tr><td>2</td><td>4</td><td>--</td><td>6</td><td>8</td></tr>
<tr><td colspan="2">$Q_1 = 3,5$</td><td>Me =5</td><td colspan="2">$Q_3 = 6,5$</td></tr>
<tr><td colspan="5"></td></tr>
<tr><td>2</td><td>4</td><td>6</td><td>8</td><td>10</td></tr>
<tr><td colspan="2">$Q_1 = 4$</td><td>Me = 6</td><td colspan="2">$Q_3 = 8$</td></tr>
</table>

Tabla 4.4. Ejemplo del método de Excel para calcular los cuartiles

En la Tabla 4.5 se presenta un resumen de los valores obtenidos cuando se aplica cada método.

Método	Datos: 2, 4, 6, 8		Datos: 2, 4, 6, 8, 10	
	Q_1	Q_3	Q_1	Q_3
Tukey	3	7	4	8
Moore McCabe	3	7	3	9
Minitab	2,5	7,5	3	9
Excel	3,5	6,5	4	8

Tabla 4.5. Resumen de los cuatro métodos utilizados para calcular los cuartiles

Entonces surgen dos preguntas: ¿cuándo se calculan los cuartiles? y ¿cuál es el método más apropiado? Para la primera pregunta, el cálculo de cuartiles tiene sentido cuando se cuenta con una población de datos relativamente grande, y para la segunda pregunta, puede usarse cualquier método ya que las diferencias entre los resultados arrojados son muy pequeñas.

4.8.2 Deciles (D_i) y Percentiles (P_i).

Los deciles son los valores de la variable que dividen a la distribución de datos en partes iguales, cada una de las cuales engloba el 10 % de los datos; en total habrá 9 deciles. Los percentiles son los valores que dividen a la distribución de datos en 100 partes iguales, cada una de las cuales engloba el 1% de las observaciones; en total habrá 99 percentiles.

4.9 Moda (M_0).

La moda, que se denotará como Mo, es el valor de la variable que más veces se repite. Cuando se tienen distribuciones de frecuencias agrupadas en intervalos y se identifica en la columna de frecuencias el valor de la distribución al que corresponde la mayor frecuencia, el respectivo intervalo se lo llamará clase modal.

En cuanto al cálculo de la moda, la determinación de la moda no tiene mayor problema si hay pocos datos y/o no están agrupados. Sin embargo, la situación es un poco más complicada si los datos vienen agrupados. En este caso, cuando los intervalos son de distinta amplitud, se define el intervalo modal, y se denota por $(L_{i-1}, L_i]$, como aquel que posee mayor densidad de frecuencia, donde $h_i = n_i/c_i$, y n_i es la frecuencia absoluta y c_i la amplitud del intervalo i. Bajo esta notación se tiene que la moda es dada por:

$$Mo = L_{i-1} + \frac{h_{i+1}}{h_{i-1}+h_{i+1}} c_i$$

En el caso de intervalos de igual longitud las densidades de las frecuencias se cambian por las respectivas frecuencias absolutas. En el caso de esta medida de tendencia central, a veces aparecen distribuciones de variables con más de una moda –llamadas bimodales, trimodales, etcétera– o incluso con una moda absoluta y otras relativas.

En cuanto a ventajas y desventajas de la moda se mencionan las siguientes:

- Su determinación es muy sencilla, cuando los datos no están agrupados y es de fácil interpretación.
- Es la única medida de posición central que puede obtenerse en las variables de tipo cualitativo de escala nominal.
- En su determinación no intervienen todos los valores de la distribución de datos.

4.10 La relación $\bar{X} > Me > Mo$.

Es interesante identificar algunas relaciones que existen entre diferentes medidas de tendencia central. Por ejemplo, si la distribución de datos es bastante simétrica, los valores de la media aritmética, la mediana y la moda, tenderán a ser iguales.

Cuando $\bar{X} > Me > Mo$, la distribución de los datos tendrá una asimetría positiva, mientras que si $\bar{X} < Me < Mo$ la distribución de datos tendrá una asimetría negativa. Además, entre media aritmética, mediana y moda, la relación ($\bar{X}$ – Mo)≅ 3($\bar{X}$ – Me), se puede verificar de manera empírica, cuando se tiene una distribución de datos convexa y moderadamente asimétrica, ya que la mediana cae entre la media aritmética y la moda, quedando aproximadamente dos veces más lejos de esta última que de la primera.

4.11 La relación $H < G < \bar{X} < Q$ para dos datos.

Resulta muy interesante y bastante ilustrativo considerar algunas construcciones geométricas que permiten comparar las medias armónica (H), geométrica (G), aritmética ($\bar{X}$) y cuadrática (Q), y concluir la desigualdad Q> $\bar{X}$ >G>H.

Primera ilustración. Inicialmente se realizará en el plano coordenado XY la construcción que se muestra en la siguiente figura, y luego se demostrará la desigualdad para dos valores positivos a y b.

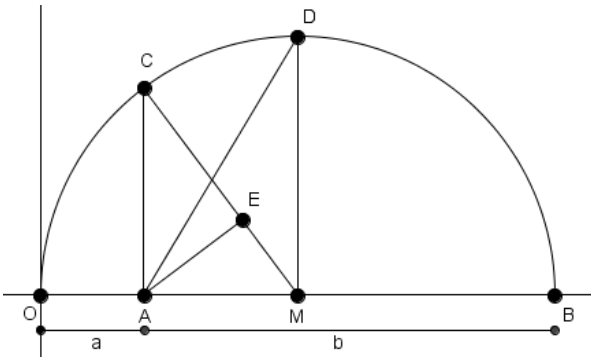


Figura 4.1.

Sobre una recta horizontal se ubican los puntos O, A y B tales que las longitudes de los segmentos OA y AB sean respectivamente a y b.

M es el punto medio del segmento OB y OM mide $\frac{a+b}{2}$.

Se traza una semicircunferencia con centro en M y radio OM.

Los segmentos CA y DM son perpendiculares al segmento OB. Los puntos C y D son puntos de la semicircunferencia.

El segmento AE es perpendicular al segmento CM.

Los puntos O, C y B son vértices de un triángulo rectángulo. En un triángulo rectángulo, la altura correspondiente a la hipotenusa es media geométrica de los segmentos en los cuales la hipotenusa es dividida por dicha altura. Aplicando este teorema, el segmento CA es media geométrica de los segmentos OA y AB, es decir, CA= $\sqrt{ab}$.

El segmento AM mide $\frac{b-a}{2}$ y el segmento DM mide $\frac{a+b}{2}$. Aplicando el teorema de Pitágoras en el triángulo AMD se obtiene que el segmento AD mide $\sqrt{\frac{a^2+b^2}{2}}$.

El segmento CM mide $\frac{a+b}{2}$ y los triángulos CAM y CEA son semejantes. De la proporción $\frac{CE}{CA}=\frac{CA}{CM}$ resulta $CE=\frac{2ab}{a+b}$.

Hasta aquí se tiene que $CE=\frac{2ab}{a+b}=H$ (Media armónica), $CA=\sqrt{ab}=G$ (Media geométrica), $CM=DM=\frac{a+b}{2}=\bar{X}$ (Media aritmética) y $AD=\sqrt{\frac{a^2+b^2}{2}}=Q$ (Media cuadrática).

Para demostrar la cadena de desigualdades, observe que en el triángulo AEC, $CE=\frac{2ab}{a+b}=H$ es un cateto y $CA=\sqrt{ab}=G$ es la hipotenusa, por lo tanto H<G. En el triángulo MAC, $CA=\sqrt{ab}=G$ es un cateto y $CM=\frac{a+b}{2}=\bar{X}$ es la hipotenusa, entonces G< $\bar{X}$. En el triángulo AMD, $DM=\frac{a+b}{2}=\bar{X}$ es un cateto y $AD=\sqrt{\frac{a^2+b^2}{2}}=Q$ es la hipotenusa, por lo tanto $\bar{X}$ <Q. Así queda demostrado que H<G< $\bar{X}$ <Q.

4.12 La relación $H_n < G_n < \overline{X_n} < Q_n$.

Si x_1, x_2, …, x_n son n números reales positivos entonces se cumple la relación

$$\frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_n}} \le \sqrt[n]{x_1x_2\ldots x_n} \le \frac{x_1+x_2+\ldots+x_n}{n} \le \sqrt{\frac{x_1^{\ 2}+x_2^{\ 2}+\ldots+x_n^{\ 2}}{n}}.$$

Para probar esta afirmación se debe tener en cuenta que si x_1, x_2, … x_n son números positivos tales que $x_1x_2\ldots x_n=1$, entonces $x_1+x_2+\ldots+x_n \ge n$.

4.13 Ejercicios (I).

Para cerrar esta sección se presenta una recopilación de ejemplos con los que se pretende ilustrar buenos y malos usos de este tipo de estadísticas.

1. Estimación del peso de un objeto.

Nueve estudiantes pesaron un objeto pequeño con un mismo instrumento en una clase de ciencias. Los pesos registrados por cada estudiante (en gramos) son: 6.2, 6.0, 6.0, 15.3, 6.1, 6.3, 6.2, 6.15 y 6.2. Los estudiantes quieren determinar con la mayor precisión posible el peso real del objeto. ¿Cuál de los siguientes métodos recomiendas usar?

a) Usar el número más común, que es 6.2
b) Usar 6.15, puesto que es el peso más preciso
c) Sumar los nueve números y dividir la suma por nueve
d) Desechar el valor 15.3, sumar los otros ocho números y dividir por ocho.

2. Promedio de niños por familia.

El comité escolar de una pequeña ciudad quiso determinar el número promedio de niños por familia en su ciudad. Dividieron el número total de niños de la ciudad por 50, que es el número total de familias. ¿Cuál de las siguientes frases debe ser cierta si el número promedio de niños por familia es 2.2?

a) La mitad de las familias de la ciudad tienen más de 2 niños.
b) En la ciudad hay más familias con 3 niños que con 2 niños.
c) Hay un total de 110 niños en la ciudad.
d) Hay 2.2 niños por adulto en la ciudad.
e) El número más común de niños en una familia es 2.

3. Media de vida en países de habla hispana.

En la Tabla 4.6 se presentan los promedios de vida de diecinueve países de habla hispana. ¿Cuál es la media de vida en países de habla hispana?

País	Promedios	País	Promedios
Cuba	78	Ecuador	71
Chile	77	El Salvador	62
Costa Rica	77	España	65
Argentina	75	Guatemala	63
Uruguay	75	Nicaragua	76
Venezuela	75	Panamá	69
México	74	Paraguay	72
Perú	71	Puerto Rico	77
Bolivia	65	Dominicana (Rep.)	100
Colombia	76		

Tabla 4.6. Media de vida en diecinueve países de habla hispana.

4. "Promedios" de bateo en el béisbol.

Los datos de la Tabla 4.9 muestran el desempeño al bate de jugadores venezolanos de béisbol que juegan en la liga Nacional y en las grandes ligas del béisbol norteamericano:

a) ¿Cuál es el "promedio" de bateo del jugador Alfonso E?
b) ¿Cuántos hits sencillos tiene que lograr Blanco H. para tener el mismo "promedio" de bateo que Escobar A?
c) ¿Es posible que Torrealba Y. con diferencia de 3 turnos al bate logre conseguir el mismo "promedio" de bateo de Pérez T? ¿Cómo?
d) Si Ordoñez M. lograra en su próximo partido conectar 3 hits en 4 turnos al bate, ¿cómo cambiaría su "promedio"?

Jugador	Liga	B	S	D	T	J	BB	AVG
Blanco H.	Americana	118	11	7	0	3	10	0.194
Escobar A.	Americana	124	18	7	1	1	12	0.241
Torrealba Y.	Americana	49	4	2	0	1	5	0.159
Alfonso E.	Nacional	157	29	7	0	1	17	
Pérez T.	Nacional	52	4	3	1	1	3	0.183
Gonzales A.	Nacional	162	16	10	0	3	20	0,204
Ordoñez M.	Americana	188	34	7	2	8	25	0,313
Olmedo R.	Nacional	1	0	0	0	0	0	0,000

Tabla 4.9. Desempeño al bate de jugadores de béisbol de Venezuela
en las ligas Nacional y Americana.

5. Notas en el colegio.

En el Colegio Distrital Luis Carlos Galán, al finalizar el año escolar la profesora de español se dispone a realizar el parcial final, Camilo le pide a la profesora sus notas y ella le muestra el siguiente reporte:

Actividad	Talleres	Cuaderno	Tareas	Parcial 1	Parcial 2	Parcial F
%	30%	10%	10%	15%	15%	20%
NOTA	3,4	2,5	3,0	2,1	3,0	

a) ¿Qué nota mínima debe sacar en el parcial final para aprobar la materia?

b) Si la profesora decide no realizar el Parcial Final y cambiar el porcentaje de los Parciales 1 y 2 cada uno al 25%. ¿Con qué nota finaliza el año?

c) Si aprueba español con 3,2. ¿Qué nota sacó en el Parcial Final?

6. Sueldos en una empresa.

En una empresa de odontología los sueldos correspondientes a los cargos son los siguientes:

Gerente General	$ 18000,000	1
Director de DPTO	$ 4'000,000	5
Jefe Inmediato	$ 3'000,000	5
Analista	$ 1'500,000	10
Auxiliares	$ 800,000	10
Servicios Generales	$ 550,000	2

El analista de nómina dice que el promedio de salario de la empresa es de $1'846,875 ¿Cómo calculó el analista de nómina el promedio de salario de la empresa?

7. ¿Conviene apostar?

Se propone un juego de dados donde el jugador participante debe lanzar tres dados de seis caras al mismo tiempo. Las reglas del juego son:

- si al lanzar los dados saca en uno de ellos un 6 gana $1.000,
- si el al lanzar los dados saca en dos ellos un 6 gana $2.000,
- si al lanzar los dados saca en los tres dados un 6 gana $3.000 y
- si no obtiene 6 en ningún dado, entonces pierde $1.500.

¿Estaría dispuesto a jugar este juego?

8. Salario de un obrero.

Un obrero eventual que trabajó ciertos meses cada año, cobró la misma cantidad de dinero cada año. Si el salario mensual fue de $80.000 en 1989, de $95.000 en 1990 y de $120.000 en 1991. ¿Cuál es el salario medio al mes del obrero?

9. Velocidad de remado.

Una persona rema río arriba a una velocidad de 25 km/h y de regreso a 40Km/h. ¿cuál es la velocidad promedio de todo el recorrido?

10. Tiempo para realizar una obra de construcción.

José puede hacer una obra en cuatro días, Juan en seis días y Andrés en doce días. ¿En cuánto tiempo pueden hacer la obra los tres juntos?

11. Crecimiento de un depósito de ahorro.

A continuación se muestra el crecimiento de un depósito de ahorro de $1000 durante cinco años, de acuerdo a las tasas de interés de 7, 8, 10, 12 y 18 por ciento para los años 1, 2, 3, 4 y 5 respectivamente.

a) ¿Cuál es el factor de crecimiento promedio?

b) ¿A qué tasa de interés corresponde el factor de crecimiento?

Año	Porcentaje de la tasa de interés	Factor de crecimiento	Ahorros al final del año ($)
1	7	1.07	1070.00
2	8	1.08	1155.60
3	10	1.10	1271.16
4	12	1.12	1423.69
5	18	1.18	1679.95

12. Salarios en una compañía de telefonía celular.

Los salarios mensuales pagados en una compañía de telefonía celular, son los siguientes:

$342.000	30 Obreros
$390.000	3 Técnicos, 1 Asistente
$392.000	35 Obreros
$440.000	1 Técnico, 1 Asistente
$442.000	2 Interventores, 2 Ingenieros
$490.000	5 Técnicos, 40 Obreros
$492.000	20 Obreros
$540.000	3 Asistente, 2 Interventores, 3 Ingenieros
$542.000	4 Interventores, 2 Directores
$590.000	2 Asistentes, 2 Ingenieros
$592.000	2 Interventores, 10 Obreros
$642.000	1 Director, 2 Asistentes

a) ¿Cuál es el salario que mejor representa el sueldo de los empleados de la empresa de telefonía celular?

b) Organiza los datos en una tabla de datos agrupados, utilizando seis intervalos de clase y determina para estos datos agrupados, la media aritmética, la mediana y la moda.

c) ¿Cuáles son los salarios medios que mejor representan los salarios de los empleados si se discriminan por tipo de empleo en la empresa?

d) Determina cual es el salario máximo que tiene el 25% del grupo de empleados con salarios más bajos, y el salario mínimo que tiene el 25% del grupo de empleados con salarios más altos.

13. ¿Cuál es la edad ideal para casarse?

Se realizó una encuesta para indagar sobre la edad en la que a las mujeres les gustaría casarse. Para ello se les preguntó a 113 mujeres, que fueron a un supermercado de Bogotá a realizar alguna compra durante el mes pasado, cuál era su opinión al respecto. Las respuestas obtenidas se organizaron en un gráfico de puntos que se presenta más adelante.

Para facilitar el estudio de los datos obtenidos en la encuesta, primero se pide que se agrupen los datos que se presentan en el gráfico de puntos, en cinco grupos de edades y con base en ésta que se responda a las siguientes preguntas:

a) ¿Entre que valores oscila la edad más común para casarse?

b) ¿Cuáles son los dos intervalos de edad que en opinión de las mujeres son apropiados para casarse?

c) ¿Cuál es el intervalo en el que para las mujeres no es muy común casarse?

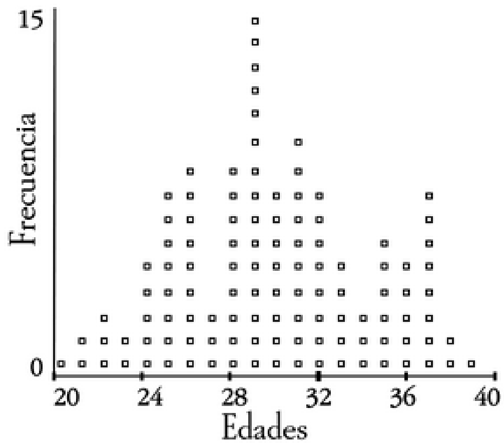


Figura 4.10. Gráfico de puntos.

14. Mediciones con un planímetro.

Un planímetro es un instrumento para medir áreas de superficies planas. Con este instrumento se tomó una muestra de cinco mediciones a las áreas de un lote de baldosas que se utilizarán en una construcción para realizar un control de calidad. La tabla siguiente muestra este conjunto de mediciones.

Baldosa	Área (cm^2)
1	402
2	398
3	405
4	406
5	400

a) Determine el error medio cuadrático.

b) De acuerdo al resultado anterior, y suponiendo que la muestra tomada es representativa de un lote aceptable de producción, ¿sería "extraño" o poco usual encontrar una baldosa que midiera 412 cm^2?

15. Interpolación para datos agrupados.

Una fábrica de bombillas realiza un proceso de control de calidad que consiste en instalar 200 bombillas en 200 portalámparas que se encuentran debidamente organizadas en varias mesas, todas son encendidas al tiempo y se registra en tiempo de duración de cada bombilla. Supóngase que el registro de las duraciones o vida útil de cada bombilla se presenta en la tabla que se muestra más adelante. Con base en la tabla se debe determinar la vida media, la moda, la mediana, los cuartiles, los deciles y los percentiles de la vida útil de las bombillas.

Duración (Horas)	Cantidad de bombillas (Fi)
[10,35)	5
[35,60)	5
[60,85)	10
[85,110)	15
[110,135)	15
[135,160)	20
[160,185)	25
[185,210)	15
[210,235)	35
[235,260)	30
[260,285)	50
[285,310)	10
[310,335)	5
[335,360]	10
Total	250

Marca de Clase Mi	Fi x Mi
22,5	113
47,5	238
72,5	725
97,5	1463
122,5	1838
147,5	2950
172,5	4313
197,5	2963
222,5	7788
247,5	7425
272,5	13625
297,5	2975
322,5	1613
347,5	3475
Total	51500

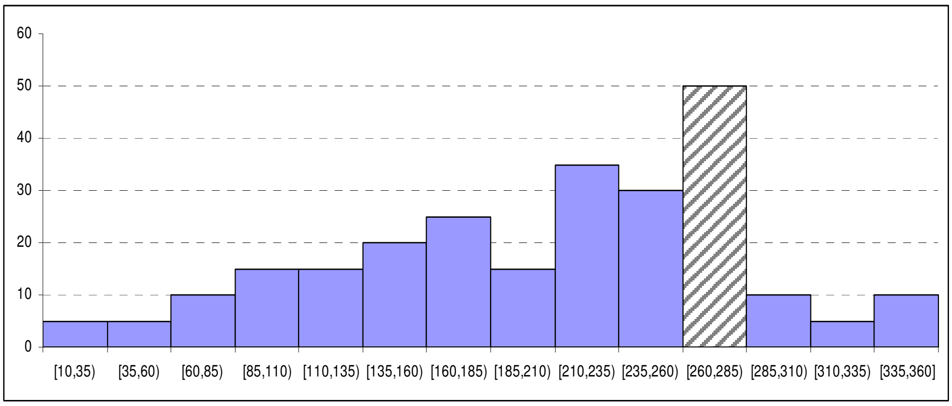


Figura 4.12.

4.14 Ejercicios (II).

1. Demostrar la fórmula de la mediana para datos agrupados suponiendo que la variable es continua.
2. Obtener una fórmula para calcular los cuartiles en datos agrupados, suponiendo que la variable es continua y que los intervalos de clase son de igual longitud.
3. Obtener una fórmula para calcular los deciles en datos agrupados, suponiendo que la variable es continua y que los intervalos de clase son de igual longitud.
4. La Universidad Nacional tiene organizadas las carreras por edificios, por ejemplo el edificio de medicina se encuentra a 27.5 m de la entrada principal, el de arte a 29.9 m y el de psicología a 28.7m. En la siguiente tabla se registra el tiempo que gastan 4 estudiantes en desplazarse de la entrada a sus respectivos edificios

Estudiante	Tiempo (minutos)	Distancia(edificio)
1	5	27.5 m
2	8	29.9 m
3	8	28.7 m
4	6	26 m

¿Cuál es el tiempo promedio que gastan las estudiantes en llegar a su respectivo edificio?

5. Verificar los diferentes cálculos de medias, medias ponderadas y medianas que aparecen reportados en el Ejemplo 3.

6. Explique por qué para obtener la velocidad promedio de un objeto que viaja con dos velocidades diferentes durante tiempos iguales, la media aritmética de las velocidades es apropiada.

7. El sábado pasado, Cristian un empleado de un kiosco de bebidas sirvió en total 50 bebidas durante la mañana de ese día. Vendió 5 bebidas de $250, 15 de $270, otras 15 de $300 y otras 15 de $350. ¿A cómo vendió en promedio cada bebida ese día?

8. En Miss Universo las cinco finalistas candidatas a la corona obtuvieron los siguientes puntajes:

Tipo de traje	Colombia	Venezuela	México	R. Dominicana	Rusia
Baño	9,40	9,20	9,00	8,90	8,85
Gala	9,83	9,40	8,90	8,88	9,00
Típico	9,48	9,60	9,20	9,80	8,50

a) Si se escogiera a la reina teniendo en cuenta el mayor puntaje obtenido. ¿Cuál sería la Reina?

b) ¿Quiénes serían la Virreina, 1ª Princesa, 2ª Princesa y 3ª Princesa?

9. Considere el siguiente juego con tres dados de seis caras: si la suma de los tres dados es 18 gana $3000, si la suma de los tres dados es 12 gana $2.000, si la suma de los tres dados es 6 gana $1.000, y si la suma es diferente de 6,12 y 18 pierde $500.

a) ¿Es conveniente jugar este juego? Justifique su respuesta en términos de una ponderación apropiada a los valores de ganancia y pérdida.

b) Construya una simulación en Excel y compare resultados teóricos y de la simulación para 100, 500 y 1.000 juegos.

10. Un estudiante de Administración de Empresas está cursando el quinto semestre, el promedio de notas de los semestres anteriores ha sido el siguiente:

Semestre	I	II	III	IV
Nota promedio	3,8	3,7	3,9	4,0

Las notas obtenidas en el quinto semestre son las siguientes:

Materia	**Créditos**	**Nota**
Mercadotecnia	3	3,8
Macroeconomía	5	4,5
Matemática Financiera	4	4,2
Presupuestos	3	4,0
Inglés	2	3,5

a) ¿Cuál es el promedio que obtuvo el estudiante en el quinto semestre?

b) ¿Qué promedio lleva el estudiante hasta el momento?

c) ¿Qué promedio debe tener en VI semestre para que su promedio general sea 4,0?

d) ¿Es posible que el promedio total hasta VI semestre sea mayor 4,5?

11. Los siguientes datos corresponden al crecimiento de un parque Automotor en la ciudad de Bogotá durante 5 años:

Años	Número de Matriculas	Factor de Crecimiento
2001	1200	1,20
2002	1440	1,25
2003	1800	1,30
2004	2340	1,20
2005	2808	1,25

¿Cuál es la tasa promedio de crecimiento del parque automotor en la ciudad?

12. Una profesora quiere cambiar la colocación de sus alumnos en clase, con la esperanza de que ello incremente el número de preguntas que hacen. En primer lugar, decide ver cuántas preguntas hacen los estudiantes con la colocación actual. El registro de número de preguntas hechas por sus ocho estudiantes durante la clase se muestra a continuación.

	Iniciales del alumno.							
	A.A.	R.F.	A.G.	J.G.	C.K.	N.K.	J.L.	A.W.
Nº de preguntas	0	5	2	22	3	2	1	3

La profesora quiere representar o resumir con un número la cantidad de preguntas hechas por estudiante. ¿Cuál de los siguientes procedimientos le recomendarías? Explica porqué.

a) Usar el número más común, que es el 2.

b) Sumar los 8 números y dividir por 8.

c) Descartar el 22, sumar los otros 7 números y dividir por 7.

d) Descartar el 0, sumar los otros 7 números y dividir por 7.

e) Otro método ¿Cuál?

13. En los últimos cinco años los suscriptores a telefonía celular han aumentado notoriamente de tal manera que el 1º de enero del año 1999 había 12.000 suscriptores y al 31 de diciembre del 2004 había 4.600.000 suscriptores. La siguiente tabla muestra los detalles:

Crecimiento de población de suscriptores				
Año	Suscriptores al inicio del año	Tasa de Crecimiento	Suscriptores al final del año	Factor de Crecimiento
1999	12.000		36.000	
2000	36.000		120.000	
2001	120.000		400.000	
2002	400.000		900.000	
2003	900.000		2.000.000	
2004	2.000.000		4.600.000	

¿Cuál es la tasa promedio de crecimiento anual y cuál es el factor promedio de crecimiento?

14. Se sabe que dos obreros A y B gastan, respectivamente, 50 y 40 minutos en remontar un par de zapatos. ¿Cuál es el tiempo promedio requerido para montar un par de zapatos?

15. Un avión recorre 3000 km. Los 1000 primeros a 700 km/h, los 1000 siguientes a 800 km/h, y los 1000 restantes a 900 km/h. ¿Cuál ha sido la velocidad media?

16. En el circuito de Indianápolis 5 competidores presentaron las siguientes estadísticas en las 5 primeras y 5 últimas vueltas:

Vuelta Nº	Cinco primeras Vueltas	Cinco últimas vueltas
	Velocidad	Velocidad
Montoya	288 km/h	312 km/h
Schumacher	300 km/h	310 km/h
Barichelo	320 km/h	290 km/h
Raikonem	290 km/h	280 km/h
Alonso	322 km/h	300 km/h

a) Calcule velocidad promedio en las cinco primeras vueltas, en las últimas cinco vueltas.

b) ¿Qué competidor tuvo el mejor promedio de velocidad teniendo en cuenta la información dada en la tabla?

17. Una población que tenía 10.000 habitantes en el año cero (2000), creció el primer año a una tasa del 5%, el segundo año creció a una tasa del 20% y el tercer año al 50%. ¿A qué tasa promedio ha crecido la población en esos tres años?

18. Los datos que se presentan en la tabla corresponden al ingreso percápita, de países de América del sur. El ingreso percápita es un indicador económico que hace referencia a todas las entradas económicas que recibe un país (en este caso en millones de dólares) dividido por el total de su población. Este indicador se considera como el ingreso medio percápita de los habitantes de cada país.

País	1995	1998
Argentina	9.728	8.030
Bolivia	2.205	1.010
Brasil	6.460	4.630
Chile	8.507	4.990
Colombia	5.861	2.470
Ecuador	3.003	1.520
Paraguay	4.312	1.760
Perú	4.180	2.440
Uruguay	8.541	6.070
Venezuela	5.706	3.530
Guayana	7.504	7.200
Surinam	2.304	5.432

a) ¿Cuál es el promedio del ingreso percápita en millones de dólares de América del sur en 1995?

b) ¿Cuál es el promedio del ingreso percápita en millones de dólares de América del sur en 1998?

c) ¿En qué año América del sur tuvo mayor ingreso percápita?

d) ¿Cuál es el promedio de ingreso percápita en millones de dólares en los dos años?

19. La cajera de una tienda va anotando los precios y las cantidades de los productos que ha adquirido un cliente. En el ticket de compra aparece esta relación:

Producto	Nº unidades	Precio por unidad
Azúcar	4	156
Aceite girasol	10	115
Leche semi-descremada	10	64
Zumo	6	75
Latas de refrescos	12	50
Botella de vino	2	139

¿Cuál será el precio superado por la mitad del precio de los productos?

20. Las tiendas High Fidelity etiquetan su mercancía 35% por encima del costo de su última adición al inventario. Hasta hace 4 meses, la grabadora de marca Mp3-Dynamic 400S costaba $300.000. Durante los últimos 4 meses High Fidelity recibió 4 embarques mensuales de esta grabadora con los siguientes costos unitarios: $275.000, $250.000, $240.000 y $225.000. ¿A qué tasa promedio mensual ha disminuido el precio de venta de High Fidelity en estos 4 meses?

21. Los datos contenidos en la siguiente tabla muestran el desempeño de 28 equipos de la Liga Nacional de fútbol Americano en 1976.

Equipos	Juegos ganados	Equipos	Juegos ganados
Washington	10	Denver	30
Minnesota	11	Detroit	6
New England	11	Green Bay	5
Oakland	13	Houston	5
Pittsburgh	10	Kansas City	5
Baltimore	11	Miami	6
Los Ángeles	10	New Orleans	4
Dallas	11	New York Giants	6
Atlanta	4	New York Jets	6
Buffalo	1	Philadephia	17
Chicago	7	San Francisco	8
Cincinnati	10	Tampa Bay	0
Cleveland	9		

¿Cuál es el número que usted escogería para representar el desempeño de los equipos de fútbol? ¿Por qué?

22. En un zoocriadero destinado a la cría de chigüiros para exportación se ha descuidado la alimentación de los animales y se ha presentado un desarrollo inesperado en estos. Se han clasificado los animales en 10 grupos, teniendo en cuenta sus pesos en kilogramos. La siguiente tabla muestra la cantidad de animales en cada categoría de pesos:

Intervalos de pesos	Número de animales
35.00 - 40.00	20
40.10 - 45.00	25
45.10 - 50.00	30
50.10 - 55.00	10
55.10 - 60.00	15
60.10 - 65.00	20
65.10 - 70.00	25
70.10 - 75.00	35
75.10 - 80.00	10
80.10 - 85.00	10

Calcule la media para estos datos agrupados y estime el valor de la mediana para estos datos agrupados.

23. En el almacén **Tornillo Loco** llegó un pedido de 25 tornillos de 3 centímetros de largo. Cuando el dueño del almacén revisó su mercancía encontró 5 tornillos con las siguientes medidas:

Tornillo	1	2	3	4	5
Medida (mm)	5	4	6	8	6

¿Cuál es el error medio cometido, prescindiendo de si este ha sido por exceso o por defecto?

24. Las siguientes fueron la causas de mortalidad de 100.000 jóvenes de Medellín, Bogotá, Cali y Barranquilla de 1989 a 1999:

10.000	mueren por tener cáncer y se suicidaron
4.500	mueren por tener enfermedades transmisibles y se suicidaron
13.500	mueren por suicidio
7.500	mueren por enfermedades transmisibles
7.890	mueren en accidentes de tránsito
5.400	mueren en accidentes de tránsito por causa de paro cardiaco
3.500	mueren por homicidios por tener enfermedades transmisibles
4.300	mueren por tumores malignos
5.900	mueren a causa de asma
19.900	mueren por homicidio
110	mueren por bronquitis
10.000	mueren accidentes de tránsito
2.500	mueren por enfermedad cerebro-vasculares de tumor maligno
5.000	mueren debido a enfermedad del sistema urinario

Identifique cuál es la causa más frecuente de mortalidad de los jóvenes de estas ciudades en las categorías relativas a: homicidios, accidente de transporte terrestre, enfermedad transmisible, tumores malignos (cáncer, leucemia, tejido linfático, etc.), enfermedad del aparato respiratorio, enfermedad cerebro vascular, agresiones auto infligidas (suicidios) y enfermedad del sistema urinario.

25. En el Colegio Cooperativo la maestra pidió a sus alumnos que con ayuda de sus padres midieran el largo de la terraza. Los miembros de la familia de David miden la longitud de la terraza en metros y encuentran los resultados que aparecen a continuación: 15.25, 12.32, 16.15, 15.25 y 11.28. ¿Cuál es el error medio cuadrático de estas mediciones?

26. Demuestre que cuando se trabaja con datos agrupados, los cuantiles se pueden calcular aplicando la fórmula $C_i(s) = L_{i-1} + \dfrac{\frac{i \cdot N}{s} - F_{i-1}}{f_i} \cdot a_i$ para i = 1, 2, …, s-1., donde i representa el i-ésimo cuantil, L_{i-1}, f_i y a_i designa el límite inferior, la frecuencia absoluta y la amplitud del intervalo, respectivamente, de la clase a la que pertenece el cuantil y F_{i-1} la frecuencia acumulada absoluta de la clase anterior a ella.

27. El conteo bacterial de cierto cultivo pasó de 1000 a 6000 en 6 días. ¿Cuál fue el promedio del incremento porcentual por día?

28. Durante cinco años sucesivos un agrónomo compró aceite diesel para sus tractores a los precios respectivos de \$125, \$200, \$350, \$430 y \$580 por galón. ¿Cuál fue el promedio del costo por galón del aceite para el periodo de los 5 años?

29. En promedio, ¿cuántos aciertos se espera que obtenga una persona cuando contesta al azar las 20 preguntas de un examen en donde cada pregunta tiene 4 opciones de respuesta, pero solo una opción es correcta?. Justifique su respuesta en términos de una ponderación apropiada para la cantidad de aciertos posibles.

30. Una caja de compensación ha clasificado a sus afiliados en 10 categorías, de acuerdo a sus salarios. Un afiliado se ubica en la categoría C(n) si su salario se encuentra en el intervalo [535.600n, 535.600 (n+1)), donde n=1,2,3,…,10. La tabla muestra la cantidad de empleados en cada categoría.

C(1)	C(2)	C(3)	C(4)	C(5)	C(6)	C(7)	C(8)	C(9)	C(10)
400	350	500	650	900	800	400	450	300	250

a) Calcule la Moda de la variable Salarios.

b) Grafique el diagrama de Cajas (Box Plot).

c) ¿Cuál es el máximo salario clasificado en el tercer decil?

d) ¿Cuál es el mínimo salario clasificado en el séptimo decil?

31. Un embarque de 15 computadoras similares que se envía a un distribuidor contiene 7 aparatos defectuosos. Una escuela escoge aleatoriamente 10 de estas computadoras y las compra. Se define la variable aleatoria X como el número de computadoras defectuosas entre las computadoras compradas. En promedio, ¿Cuántas computadoras defectuosas se espera que lleve el comprador?

32. Un juego de apuestas consiste en lanzar seis veces un dado normal y apostar cierta cantidad de dinero a un resultado, de tal manera que si el número apostado resulta n veces, entonces gana 500n pesos, y si no resulta pierde 2000 pesos. Un jugador apuesta al número "6". Se define la variable aleatoria como el beneficio del jugador en 6 lanzamientos de un dado. Use la media ponderada para determinar la ganancia esperada del jugador.

33. Un objeto de tiro al blanco está formado por 5 círculos concéntricos de radios 10 cm, 20 cm, 30 cm, 40 cm y 50 cm. Un hombre que dispara al blanco recibe 50 puntos, 40 puntos, 30 puntos, 20 puntos o 10 puntos, según pegue en la zona 1 (círculo pequeño), zona 2, zona 3, zona 4 o zona 5 (anillos circulares). La probabilidad de que el disparo haga contacto con cualquiera de las 5 zonas del blanco es 1/3, y la probabilidad de no dar en el blanco es 2/3. Si X se define como el puntaje en un disparo, ¿Cuál el puntaje esperado?

34. Anatoly recibe 5 cartas de una baraja francesa a la que le faltan 4 cartas de corazones y 4 cartas de diamantes, mientras que Boris recibe 5 cartas de una baraja francesa completa. ¿Quién tiene mayor expectativa de obtener corazones?

35. La empresa Cinascar se ha posicionado en el país por la venta de vehículos a precios cómodos. La empresa empezó en el 2004 vendiendo 1200 vehículos, en los años siguientes hasta el 2009, ha vendido 1440, 1800, 2340, 2808 y 3510 vehículos respectivamente. ¿Cuál es la tasa promedio de crecimiento de las ventas de estos últimos 5 años?

36. Un juego consiste en lanzar 4 dados distinguibles y apostarle a la aparición del número 6. Si éste número no sale el apostador pierde 100 pesos. Si éste le aparece 1, 2, 3 o 4 veces la ganancia será de 100, 200, 300 o 400 pesos respectivamente. ¿Cuál es la ganancia media del jugador, en un lanzamiento de los dados?

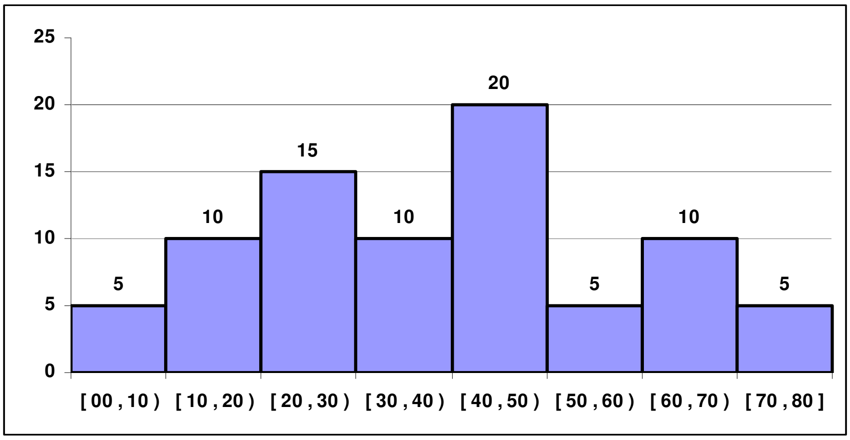
37. Use la medida ponderada para determinar la favorabilidad del siguiente juego: Se lanzan tres dados normales: si la suma de los tres dados es 18 gana $3000, si la suma de los tres dados es 12 gana $2.000, si la suma de los tres dados es 6 gana $1.000, y si la suma es diferente de 6,12 y 18 pierde $500. ¿Es conveniente jugar este juego? Justifique su respuesta en términos de una ponderación apropiada a los valores de ganancia y pérdida.

38. La serie final de un campeonato de béisbol fue disputada por los equipos A y B, cada uno con 20 jugadores. Al final de la serie se contabilizaron los batazos de hit conectados por los jugadores titulares de cada equipo y se llegó a la distribución que se muestra en la tabla, con la cual se puede estimar información más precisa de la variable "Cantidad de Hits conectados por jugador".

Hits	Jugadores
[100,120)	4
[120,140)	8
[140,160)	12
[160,180)	6
[180,200)	8
[200,220]	2

a) Calcule el promedio de la cantidad de hits conectados por jugador.
b) Calcule la mediana de la cantidad de hits conectados por jugador.
c) Calcule la moda de la cantidad de hits conectados por jugador.

39. El siguiente diagrama representa la distribución de frecuencias de los valores de una variable continua X.



a) Calcule el promedio aritmético, la mediana y la moda de la variable X.
b) Calcule el cuartil 3, el decil 7 y el percentil 65.

40. Un examen de Cálculo se aplicó a los cuatro grupos de grado 11 de una institución. En la siguiente tabla se presenta la cantidad de estudiantes con notas en cada rango de notas para cada uno de los cuatro grupos. Calcular para cada grupo la media, la mediana, la moda, el cuartel 1 y el decil 7.

Notas	Grupo A	Grupo B	Grupo C	Grupo D
[0.0 – 0.5)	8	14	2	12
[0.5 – 1.0)	6	6	4	14
[1.0 – 1.5)	8	4	6	8
[1.5 – 2.0)	10	8	10	6
[2.0 – 2.5)	12	4	12	5
[2.5 – 3.0)	8	8	10	5
[3.0 – 3.5)	8	8	6	6
[3.5 – 4.0)	6	6	4	8
[4.0 – 4.5)	14	10	2	14
[4.5 – 5.0]	8	6	14	12

41. La siguiente tabla presenta las calificaciones en una prueba específica de Matemáticas de un grupo de 150 aspirantes que han aprobado previamente una prueba de potencialidad pedagógica.

5,0	5,3	5,6	5,9	6,2	6,5	6,8	7,1	7,4	7,7	8,0	8,3	8,6	8,9	9,2
17,0	17,5	18,0	18,5	19,0	19,5	20,0	20,5	21,0	21,5	22,0	22,5	23,0	23,5	24,0
25,0	25,6	26,2	26,8	27,4	28,0	28,6	29,2	29,8	30,4	31,0	31,6	32,2	32,8	33,4
35,0	35,7	36,4	37,1	37,8	38,5	39,2	39,9	40,6	41,3	42,0	42,7	43,4	44,1	44,8
45,0	45,8	46,6	47,4	48,2	49,0	49,8	50,6	51,4	52,2	53,0	53,8	54,6	55,4	56,2
57,0	57,9	58,8	59,7	60,6	61,5	62,4	63,3	64,2	65,1	66,0	66,9	67,8	68,7	69,6
70,0	70,2	70,4	70,6	70,8	71,0	71,2	71,4	71,6	71,8	72,0	72,2	72,4	72,6	72,8
73,0	73,1	73,2	73,3	73,4	73,5	73,6	73,7	73,8	73,9	74,0	74,1	74,2	74,3	74,4
75,0	76,0	77,0	78,0	79,0	80,0	81,0	82,0	83,0	84,0	85,0	86,0	87,0	88,0	89,0

a) Los aspirantes cuyas calificaciones se ubiquen en el primer cuartil son retirados del proceso de admisión. ¿Cuál es la menor calificación entre los estudiantes que siguen en el proceso?

b) Los estudiantes cuyas calificaciones se encuentren en los dos últimos deciles no presentan entrevista y ya quedan admitidos. ¿Cuál debe ser la mínima calificación para quedar en este selecto grupo?

42. En la siguiente tabla se encuentran clasificados los puntajes de la prueba de potencialidad pedagógica de los aspirantes a un programa de licenciatura de una universidad oficial.

a) Use el método de interpolación lineal en frecuencias acumuladas para calcular los percentiles P_{35} y P_{65}, y los deciles D_3 y D_7.

b) Use la fórmula $Q_k(s) = L_{k-1} + \left(\dfrac{\frac{kN}{s} - F_{k-1}}{f_k}\right)L$, (k=1,2,…,s-1) para calcular los percentiles P_{35} y P_{65}, y los deciles D_3 y D_7.

Donde:

k es k-ésimo cuantil;

L_{k-1} es Límite inferior de la clase cuantílica;

f_k es Frecuencia absoluta de intervalo cuantílico;

Fk-1 es Frecuencia acumulada hasta el intervalo anterior al intervalo cuantílico;

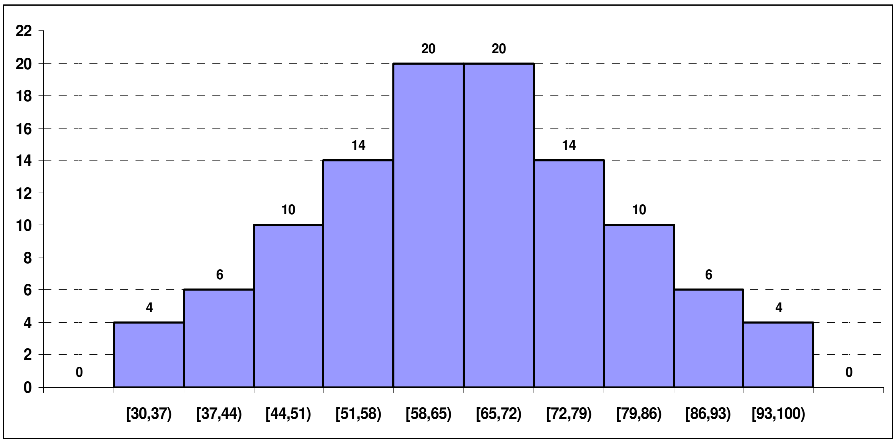
L: Longitud de clases; s= 4 (cuartiles), 10 (deciles), 100 (percentiles).

Calificaciones	Cantidad de Aspirantes
[0.0 – 0.5)	45
[0.5 – 1.0)	50
[1.0 – 1.5)	20
[1.5 – 2.0)	35
[2.0 – 2.5)	40
[2.5 – 3.0)	20
[3.0 – 3.5)	40
[3.5 – 4.0)	30
[4.0 – 4.5)	15
[4.5 – 5.0)	15

Calificaciones	Cantidad de Aspirantes
[5.0 – 5.5)	45
[5.5 – 6.0)	55
[6.0 – 6.5)	50
[6.5 – 7.0)	35
[7.0 – 7.5)	25
[7.5 – 8.0)	25
[8.0 – 8.5)	40
[8.5 – 9.0)	30
[9.0 – 9.5)	0
[9.5 – 10.)	0

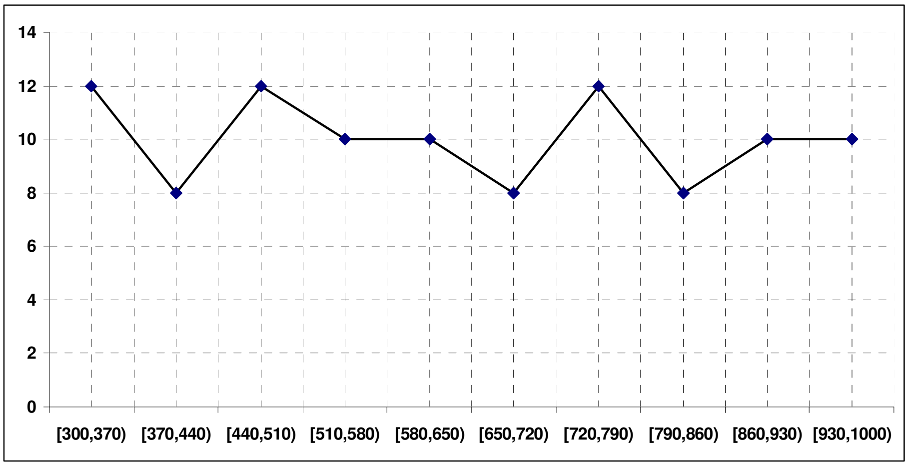
43. El histograma de la figura representa los pesos en kilogramos de 104 deportistas que representan a Colombia en unos juegos panamericanos.

a) Calcular la moda de los pesos.
b) Calcular los deciles de los pesos.



44. El polígono de frecuencias de la figura representa los salarios en euros de 100 empleados de una empresa de correos.

a) Calcular el promedio, la mediana y la moda de los salarios.
b) Calcular los cuartiles de los salarios y grafique el diagrama de cajas.



45. X es una variable continua que toma N valores distribuidos en n intervalos de longitud L, la frecuencia de la clase modal es fm y las frecuencias de las clases anterior y posterior a la clase modal son fa y fp respectivamente, los límites de las tres clases consideradas son La, Li, Ls y Lp. (Ver figura).

a) Demuestre, que la moda de datos agrupados de la variable X es igual a $Mo = Li + \left(\dfrac{Dp}{Da + Dp}\right)L$, donde Da=fm-fa y Dp=fm-fp.

b) Demuestre que la Moda es la abscisa del vértice de la parábola que pasa por los puntos $\left(\dfrac{La + Li}{2}, fa\right)$, $\left(\dfrac{Li + Ls}{2}, fm\right)$ y $\left(\dfrac{Ls + Lp}{2}, fp\right)$.