Identificación del Objeto de aprendizaje

Fecha	Julio 2013
Asignatura	Estadística y Probabilidad 1
Unidad	Unidad I. Estadística Descriptiva
Tiempo disponible	24 horas
Aprendizajes	• Identifica las variables como atributos de interés de una población o muestra. • Comprende que los datos constituyen los valores que toma una variable • Identifica variables cualitativas y cuantitativas. • Valora la importancia de la recopilación de datos en el proceso de una investigación. • Construye tablas de distribución de frecuencias para representar el comportamiento de variables cualitativas y variables cuantitativas. • Interpreta tablas para describir el comportamiento de un conjunto de datos. • Construye histogramas, polígonos de frecuencias, ojivas, gráficas de barras, circulares y de caja. • Interpreta gráficas para describir el comportamiento de un conjunto de datos. • Conoce las propiedades de las medidas de tendencia central. • Calcula la media aritmética, la mediana y la moda para datos agrupados y no agrupados. • Argumenta la elección de una medida de tendencia central para describir el comportamiento de un conjunto de datos. • Conoce el concepto de dispersión en la descripción de un conjunto de datos. • Calcula la desviación estándar y la varianza, y comprende sus significados. • Calcula el coeficiente de variación y comprende su significado. • Calcula las medidas de posición y comprende su significado.
Tema	1. Variable y recopilación de datos. 1.1. Medidas de tendencia central. • Media aritmética. • Mediana. • Moda. 1.2. Medidas de dispersión y de posición. • Desviación estándar. • Varianza. • Coeficiente de variación. • Cuantiles. 2. Tablas de distribución de frecuencias.

	3. Representaciones gráficas. • Histogramas. • Polígonos de frecuencias. • Ojivas. • Gráfica de barras. • Gráfica circular. • Gráfica de caja. 4. Medidas de tendencia central para datos agrupados. • Media aritmética. • Mediana. • Moda. 5. Medidas de dispersión y de posición para datos agrupados. • Desviación estándar. • Varianza. • Coeficiente de variación. • Cuantiles.
Palabras claves	Población, muestra, variable cualitativa o cuantitativa, tabla de datos, histogramas, polígonos de frecuencias, ojivas, gráficas de barras, circulares y de caja, medidas de tendencia central; media aritmética, la mediana y la moda para datos agrupados y no agrupados, medidas de dispersión; absoluta; desviación estándar y la varianza y relativa; coeficiente de variación. Medidas de posición; Cuartiles, deciles y porcentiles.
Autor	Tomás Espinosa Martínez.

Objetivo (para el profesor)

Trabajar con datos recopilados por los alumnos, como deporte preferido, número de hermanos, peso, estatura, longitud de los pies, el tiempo de reacción, el sabor del yogurt que consumen, el modelo de la calculadora (o marca)que ellos tienen. El tiempo que tardan en trasladarse de su casa a la escuela, el tiempo de retraso en la llegada a la clase.Se sugiere también *t*rabajar con material lúdico, por ejemplo; dominó, dados, pirinola de seis caras con letreros, y también con monedas donde es posible obtener valores ya sea numéricos o bien de atributos. La finalidad es que el comportamiento de dichos datos les resulte significativo y sea posible explicarles a los alumnos la diferencia e importancia de los conceptos: Variabilidad, variable aleatoria, variable; cualitativa y cuantitativa, medidas centrales y de dispersión; absoluta y relativa, así como mostrar las diferencias que existen en las construcciones de tablas de datos de variables cualitativas y cuantitativas , el siguiente paso es la construcción de las gráficas; histogramas, polígonos de frecuencias, ojivas, gráficas de barras, circulares y de caja, de su análisis e interpretación. Finalmente pasar al cálculo de las medidas de tendencia central; media aritmética, la mediana y la moda para datos agrupados, medidas de dispersión; absoluta; desviación estándar y la varianza y relativa; coeficiente de variación para que el estudiante realice el cálculo y comprenda las diferencias entre las medidas de dispersión, de posición y el	**Objetivo (para el profesor)**

coeficiente de variación.	

Índice de navegación del Objeto de aprendizaje

1. INTRODUCCIÓN.

Cuando se realizan mediciones o se dispone de muchas medidas, que por lo general son distintas, se presenta el problema de cuál de ellos representa mejor los datos, ¿Cuál se debe elegir, ¿Cómo tomarlos en cuenta a todos los datos?, ¿de qué manera seleccionar la mejor o al más representativo de ese conjunto de datos?, los métodos que sirven para determinarlo representa la posibilidad de sintetizar el grupo de datos en un valor representativo, también llamado valor central.

Por otra parte las medidas de dispersión nos dicen hasta qué punto está o estas medidas de tendencia central son representativas como síntesis de la información. Así un valor muy grande o muy pequeño será un indicador de la exactitud y de la calidad en el procedimiento de medición, las medidas de dispersión cuantifican la separación de los datos con respecto a la medida central o de dispersión, la variabilidad de los valores de la distribución respecto al valor central. Se puede Distinguir entre medidas de dispersión absolutas, que no son comparables entre diferentes muestras y las relativas que nos permitirán comparar varias muestras.

Dada la tabla de la distribución de frecuencias, es posible hallar medidas de tendencia central, así como también medidas de dispersión. Considera los siguientes datos, organízalos, y agrupados en categorías, de acuerdo a lo indicado en la sección 1.2.3 (tablas de distribución de frecuencias).

1.1 ALGUNAS MEDIDAS DE TENDENCIA CENTRAL Y DISPERSION PARA DATOS NO AGRUPADOS.

Problema 1.

Considere el siguiente situación; al finalizar el semestre, un profesor del Colegio CCH da a escoger a uno de sus estudiantes con qué medida central; media aritmética, mediana, moda, media armónica o media geométrica desea ser evaluado, para determinar su calificación final. Las calificaciones parciales del estudiante correspondientes a cinco evaluaciones realizadas para evaluar cinco unidades del programa, tienen el mismo peso; 4, 3, 6, 6, 9. ¿Qué medida central conviene al estudiante?

Las medidas centrales, mencionadas en el problema son: medida central; media aritmética; M_A, mediana; M_E, moda; M_O, media armónica; M_R y media geométrica; M_G. La manera de calcularlas es a través de:

TABLA 1.1 Algunas medidas centrales

media aritmética: M_A $M_A = \dfrac{\sum_{i=1}^{n} x_i}{n}$	media armónica: M_H $M_H = \dfrac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \dfrac{n}{\frac{1}{x_1} + \frac{1}{x_2} ... + \frac{1}{x_n}}$
Mediana: M_E $\begin{cases} si\ es\ impar: M_E = x_{\frac{(n+1)}{2}} \\ si\ es\ par: M_E = \dfrac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \end{cases}$	media ponderada: M_{AP} $M_{AP} = \dfrac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i}$
media geométrica: M_G $M_G = \sqrt[n]{\underset{i=1}{\overset{n}{\pi}} x_i} = \sqrt[n]{x_1 \cdot x_2 ..}$	Moda: M_O : “Medida más frecuente”

Realizando los cálculos:

MEDIA ARITMÉTICA		$M_A = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{4+3+6+6+9}{5} = 5.6$
MEDIANA	$\begin{cases} si\ es\ impar: M_E = x_{\frac{(n+1)}{2}} \\ si\ es\ par: M_E = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \end{cases}$	acomodando datos en forma creciente: 3,4,6,6,9, el dato en la posición: $M_E = x_{\frac{(n+1)}{2}} = x_{\frac{(5+1)}{2}} = x_3$, es la mediana es decir $M_E = 6$
MODA	M_O : “Medida más frecuente” , El valor con mayor frecuencia es el 6, entonces: $M_O = 6$	
MEDIA ARMÓNICA	$M_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} ... + \frac{1}{x_n}} = \frac{5}{\frac{1}{3} + \frac{1}{4} + \frac{1}{6} + \frac{1}{6} + \frac{1}{9}} \approx 4.8648$	
MEDIA GEOMÉTRICA	$M_G = \sqrt[n]{\pi_{i=1}^{n} x_i} = \sqrt[n]{x_1 \cdot x_2 ... x_n} = \sqrt[5]{3 \cdot 4 \cdot 6 \cdot 6 \cdot 9} \approx 5.2233$	

ACTIVIDAD 1

Considere la situación académica de otro estudiante del CCH, que tiene calificaciones parciales: 9, 7, 4, 9, si se diera a escoger la medida central: media aritmética, mediana, moda, media armónica y media geométrica para obtener su calificación final,

1. ¿Cuál sería lamedida central que proporciona una calificación másalta?
2. ¿Cuál sería la medida central menos conveniente al estudiante como calificación?

Solución

Media aritmética	....	$M_A = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{4+\ 7+\ 9+\ 9}{4} = \frac{29}{4} = 7.25$
mediana	$\begin{cases} si\ es\ impar: M_E = x_{\frac{(n+1)}{2}} \\ si\ es\ par: M_E = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \end{cases}$	acomodando datos en forma creciente: 4,7,9,9, por ser par el número de datos, seleccionamos los datos que ocupan el ordinal segundo y tercero; 7 y 9. $M_E = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{7+9}{2}$. La mediana de las calificaciones es $M_E = 8$
moda	M_O : “Medida más frecuente” ,	El valor con mayor frecuencia es el 9, entonces: $M_O = 9$
media armónica		$M_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} \dots + \frac{1}{x_n}} =$ $= \frac{4}{\frac{1}{4} + \frac{1}{7} + \frac{1}{9} + \frac{1}{9}} \approx 6.5032$
media geométrica		$M_G = \sqrt[n]{\pi_{i=1}^{n} x_i} = \sqrt[n]{x_1 \cdot x_2 \dots x_n} =$ $= \sqrt[4]{4 \cdot 7 \cdot 9 \cdot 9} \approx 6.9009$

La medida central que proporciona una calificación más alta es: LA MODA	
La medida central menos conveniente al estudiante como calificación es: LA MEDIA GEOMÉTRICA	

ACTIVIDAD 2

Lea con cuidado los enunciados de las columnas y relacione de manera correcta ambas columnas.

Solución

a. Una empresa cotiza en la bolsa de valores BMV, las variaciones en la cotización de sus acciones, pérdidas o ganancias mensuales en cinco meses consecutivos es: 4, -2,-5, 3,-1, la medida central que no es posible calcular es:

b. Un vendedor de automóviles vendió las siguientes cantidades de autos Modelos 2013, en siete días: 3, 2, 1, 1, 0, 2,1, la medida central que no es posible calcular es:

c. si se consideran cuantas veces aparece cada dato y se hace el cálculo el resultado concuerda con la media Aritmética.

d. se considera que es la medida central más justa y equitativa

(…..) Media armónica
"no es posible dividir entre cero"

(…..) Moda
"Puede existir más de un Valor"

(…..) Mediana

"números muy Negativos o positivos grandes modifican fuertemente su valo

(…..) Media aritmética
"al cancelarse los valores extremos , los datos intermedios, no se ven afectados por valores extremos"

e. Es la medida central que se ve afectada por los valores extremos.

f. Las calificaciones finales de un grupo de estudiantes, son : 5, 6, 6, 6, 6, 6, 7 ,7,8,8,8,9,9, 9,10,10, la medida central que no es posible determinar de manera única es:

(….) Media ponderada
“su expresión algebraica se deduce de la media” aritmética”

(….) Media Geométrica
“no se pueden calcular raíces de números negativos”

Solución

a. Una empresa cotiza en la bolsa de valores BMV, las variaciones en la cotización de sus acciones, pérdidas o ganancias mensuales en cinco meses consecutivos es: 4, -2,-5, 3,-1, la medida central que no es posible calcular es:

b. Un vendedor de automóviles vendió las siguientes cantidades de autos Modelos 2013, en siete días: 3, 2, 1, 1, 0, 2,1, la medida central que no es posible calcular es:

c. si se consideran cuantas veces aparece cada dato y se hace el cálculo el resultado concuerda con la media Aritmética.

d. se considera que es la medida central más justa y equitativa

(….. b) Media armónica
“no es posible dividir entre cero”

(…..f) Moda
“Puede existir más de un Valor”

(….. d) Mediana

“números muy Negativos o positivos grandes modifican fuertemente su valo

(…..e) Media aritmética
“al cancelarse los valores extremos , los datos intermedios, no se

	ven afectados por valores extremos”
e. Es la medida central que se ve afectada por los valores extremos.	(…. c) Media ponderada “**su expresión algebraica se deduce de la media” aritmética”**
f. Las calificaciones finales de un grupo de estudiantes, son : 5, 6, 6, 6, 6, 6, 7 ,7,8,8,8,9,9, 9,10,10, la medida central que no es posible determinar de manera única es:	(….a) Media Geométrica ***“no se pueden calcular raíces de números negativos”***

Existen varias situaciones por las cuales se prefiere trabajar con la media aritmética, entre los que podemos citar:

a. Realizar el cálculo es sencillo y en él intervienen los valores de todos los datos, cosa que no tiene la mediana.
b. su valor es único para una serie de datos dada, a diferencia de la moda, que puede tener uno o más valores.

c. No tiene ningún problema si alguno de los datos es igual a cero, porque no se indetermina o bien si los datos son negativos no presenta inconsistencias (raíces de números negativos), por lo que es mejor que la media armónica y la media geométrica.

d. es apropiada para comparar poblaciones, porque es posible acompañarla de una medida de dispersión.

e. Se interpreta geométricamente como el "punto de equilibrio" o "centro de masas" del conjunto de datos, ya que tiene la propiedad de equilibrar las desviaciones de los datos respecto de su propio valor.

f. Si se seleccionaran diversos valores posibles de ella el, la media aritmética es la única que minimiza las desviaciones cuadráticas de los datos respecto de cualquier valor prefijado. Esta *propiedad permite interpretar los parámetros de dispersión más importantes: la varianza y la desviación estándar...*

g. Sin embargo también presenta desventajas: su valor se ve muy afectado por la falta de homogeneidad de los datos, por ejemplo valores extremos, por ejemplo en el caso de los salarios de los obreros y los ejecutivos, supóngase cinco obreros devengan un salario de 100 pesos diarios y u ejecutivo de 1000 pesos, el salario promedio de los cinco obreros seria 100 pesos, pero si se incluye al ejecutivo, el salario promedio de las seis personas seria 300 pesos.

ACTIVIDAD 3

Se realiza una evaluación para contratar a personal de una empresa bancaria, se de realizar tres pruebas: $A_{1,}$ A_2 y A_3, la primera es un examen de conocimientos, la segunda es una evaluación sobre su capacidad asertiva en la toma de decisiones y la tercera sobre manejo de software. La primera evaluación A1 vale tres veces más que la tercera y la segunda vale el doble de la tercera. Dos aspirantes C_1, C_2, C_3 y C_4 deben ser evaluados, y solo puede contratarse a uno de ellos. Si se utiliza la media aritmética como promedio.

a. ¿Qué puntuación obtuvo cada uno?

b. ¿Qué puntuación se obtuvo en cada una de las evaluaciones?

c. ¿Qué evaluación fue más difícil para los aspirantes?

d. c. ¿Qué evaluación fue más fácil para los aspirantes?

Solución.

Puesto que la primera evaluación vale tres veces más que la tercera y la segunda

vale el doble de la tercera. Entonces:

examen	**Peso**	**peso relativo**
examen 1	**= 3 veces examen**	**3/6**
examen 2	**= 2 veces examen 3**	**2/6**
examen 3	**= 1 vez examen 3**	**1/6**

el modelo para la evaluación tomando en cuenta las tres calificaciones:

$$C_f = A_1 \cdot (peso\ de\ A_1) + A_2 \cdot (peso\ de\ A_2) + A_3 \cdot (peso\ de\ A_3) =$$

$$C_f = A_1 \cdot \frac{3}{6} + A_2 \cdot \frac{2}{6} + A_3 \cdot \frac{1}{6} = \frac{3A_1 + 2A_2 + A_3}{6}$$

las calificaciones de los aspirantes C_1, C_2, C_3 y C_4:

$$C_f = \frac{3A_1 + 2A_2 + A_3}{6}$$

$$C_{f1} = \frac{3(7.3) + 2(5.2) + (5.9)}{6} = 6.3667$$

$$C_{f2} = \frac{3(4.5) + 2(6.3) + (9.9)}{6} = 6.0000$$

$$C_{f3} = \frac{3(5.5) + 2(6.6) + (8.4)}{6} = 6.3500$$

$$C_{f4} = \frac{3(5.5) + 2(5.8) + (9.5)}{6} = 6.2667$$

	Evaluación A_1	**Evaluación A_2**	**Evaluación A_3**	***<u>Puntuación</u>***
Aspirante C_1	**7.3**	**5.2**	**5.9**	**<u>6.3667</u>** ***Calificación más alta***
Aspirante C_2	**4.5**	**6.3**	**9.9**	**6.000**
Aspirante C_3	**5.5**	**6.6**	**8.4**	**6.3500**
Aspirante C_4	**5.5**	**5.8**	**9.5**	**6.2667**
	<u>5.700</u>	***<u>5.975</u>***	***<u>8.425</u>***	
	<u>Más difícil</u>		***<u>Más fácil</u>***	

LA MEDIA ARITMÉTICA PONDERADA.

En determinados problemas se necesita calcular la media aritmética de un conjunto de datos, en ocasiones observamos que los datos numéricos se repiten muchas veces, por ejemplo para este conjunto: $3,3,3,3,3,4,4,4,4,4,4,4,5,5,5,5,5$ Para hallar la media aritmética: $M_A = \dfrac{3+3+3+3+3+4+4+4+4+4+4+4+5+5+5+5+5}{17}$, se puede observar que el dato número "3" , se repite cinco veces, el dato con el número "4" se repite siete veces y el dato con valor "5" se repite cinco veces, la media aritmética , se podría escribir como:

$$M_A = \frac{3+3+3+3+3+4+4+4+4+4+4+4+5+5+5+5+5}{17} = \frac{3\cdot(5)+4\cdot(7)+5\cdot(5)}{(5)+(7)+(5)}$$

Las veces que se repite un dato se le llama frecuencia o peso y se simboliza como: f_i

En general; si se consideran los datos, organizados de manera creciente o decreciente para poder observar las regularidades en su repetición (frecuencias) , por ejemplo para los datos: $x_1, x_1 ..., x_1; x_2, x_2 + ..., x_2; ...; x_k, x_k, ..., x_k$, cuyas frecuencias son: $f_1, f_2, ..., f_k$, la media aritmética :

$$M_A = \frac{\overbrace{x_1 + x_1 \ldots + x_1}^{f_1} + \overbrace{x_2 + x_2 + \ldots + x_2}^{f_2} + \ldots + \overbrace{x_k + x_k + \ldots + x_k}^{f_k}}{f_1 + f_2 + \ldots + f_k}$$, de manera abreviada:

$$M_A = \frac{\overbrace{x_1 + x_1 \ldots + x_1}^{f_1} + \overbrace{x_2 + x_2 + \ldots + x_2}^{f_2} + \ldots + \overbrace{x_k + x_k + \ldots + x_k}^{f_k}}{f_1 + f_2 + \ldots + f_k} = \frac{\sum_i^k x_i \cdot f_i}{\sum_i^k x_i}$$, la expresión equivalente para calcular la media aritmética y se le llama media ponderada. $M_{Ap} = \frac{\sum_i^l x_i \cdot f_i}{\sum_i^l x_i}$, la expresión se observa complicada, pero no lo es, si se comprende cómo utilizarla. Se puede verificar que para los datos: $3,3,3,3,3,4,4,4,4,4,4,4,5,5,5,5,5$, la media aritmética y la media ponderada proporcionan el mismo resultado.

Tabla 1.2 media ponderada

media aritmética ponderad M_{AP}	$M_{AP} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{3\cdot(5)+4\cdot(7)+5\cdot(5)}{(5)+(7)+(5)} = \frac{68}{17} = 4$
media aritmética M_A,	$M_A = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{3+3+3+3+3+4+4+4+4+4+4+4+5+5+5+5+5}{17} =$ $= \frac{68}{17} = 4$

ACTIVIDAD 4

Se desea determinar el número promedio a través de refrescos consumidos por semana por una persona durante 10 semanas (Sn).

S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
4	5	5	5	5	6	6	6	6	6

Determine la media aritmética ponderada M_{AP} y compare los resultados con media aritmética M_A, determine cuales afirmaciones son correctas (**V**) o cuales son falsas (**F**).

a. $M_{AP} > M_A$ ________________

b. $M_A > M_{AP}$ ____________________

c. $M_{AP} = M_A$ _________________

d. $M_{AP} < M_A$ ______________________

e. $M_A < M_{AP}$ _________________

Solución.

a. $M_{AP} > M_A$ _______F__________

b. $M_A > M_{AP}$ ___________F__________

c. $M_{AP} = M_A$ _______V___________

d. $M_{AP} < M_A$ ___________F___________

e. $M_A < M_{AP}$ _______F___________

Se observa que el resultado es el mismo, es decir: $M_{AP} = M_A$***, entonces solo la proposición del inciso el inciso c , es verdadera y todas las demás son falsas.***

ACTIVIDAD 5

Determine el valor de la media aritmética y de la media aritmética ponderada, de los datos: 4, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6.

Solución.

media aritmética ponderada M_{AP}	$M_{AP} = \frac{\sum_{i=1}^{n} x_i f_i}{\sum_{i=1}^{n} f_i} = \frac{4(1)+5(4)+6(5)}{10} = \frac{54}{10} = 5.4$
media aritmética M_A,	$M_A = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{4+\ 5+5+5+5+6+6+6+6+6}{5} = \frac{54}{10} = 5.4$

Problema 2.

Considere el siguiente situación; se busca otorgar una beca al mejor estudiante, de la materia de probabilidad y estadística I, existen cinco candidatos; A, B, C, D y E pero por falta de presupuesto solo se otorgara a uno (al mejor estudiante). Por simplicidad suponga que se evalúan de acuerdo a las calificaciones parciales del semestre que son tres. La información se presenta en la tabla abajo. Si el promedio se obtiene con ayuda de la media aritmética, determine ¿Cuál es el mejor estudiante?

Tabla 1.3 calificaciones de alumnos

	Primera evaluación	Segunda evaluación	Tercera evaluación
Estudiante A	8.0	9.0	10.0
Estudiante	10.0	9.0	8.0

B			
Estudiante C	9.0	9.0	9.0
Estudiante D	10.0	8.0	9.0
Estudiante E	10.0	9.0	7.0

Solución.

	Media aritmética
Estudiante A	$M_A = \frac{8+9+10}{3} = \frac{27}{3} = 9$
Estudiante B	$M_A = \frac{10+9+8}{3} = \frac{27}{3} = 9$
Estudiante C	$M_A = \frac{9+9+9}{3} = \frac{27}{3} = 9$
Estudiante D	$M_A = \frac{10+8+9}{3} = \frac{27}{3} = 9$
Estudiante E	$M_A = \frac{10+9+7}{3} = \frac{26}{3} = 8.66$

Podría ser A , B, C o D

ACTIVIDAD 6

Con los datos de la tabla anterior complete los enunciados.

1. El estudiante que tiene menor promedio es el estudiante: ______

2. los demás Estudiantes tienen media aritmética igual son:__________

3. ¿El mejor estudiante, atendiendo sus calificaciones es? ______________

Solución.

Con los datos de la tabla anterior complete los enunciados.

1. El estudiante que tiene menor promedio es el estudiante: ______**E**

2. los demás Estudiantes tienen media aritmética igual son:**A,B,C y D**__

3. ¿El mejor estudiante, atendiendo sus calificaciones es? **“No es posible determinarlo”**

Es difícil discernir cual es el mejor, solo con la medida central. Quedando como finalistas los candidatos: A, C, D y E.

Algunos criterios que se podrían considerar no son imparciales y dependerían posiblemente de la cuestión sentimental por ejemplo, se podría decir que el mejor es el estudiante A, porque empezó desde una calificación de ocho y fue mejorando en el trascurso del semestre.

Sin embargo desde el punto de vista de la estadística, el mejor es el estudiante C, ¿Por qué?, ¿en qué criterio justifica esta apreciación?

1.2 ALGUNAS MEDIDAS DE DISPERSIÓN. LA VARIANZA Y LA DESVIACIÓN ESTÁNDAR

Imagine que se trata de máquinas que de manera automática llenan algún producto en bolsas de 9 kilogramos, por ejemplo arroz, la capacidad que tienen las bolsas es de solo 9 kilos, si la maquina vacía en determinado momento, 9 o 8 o 10 kilos, algunas veces quedara espacio en las bolsas, en otras el producto se derrama en el piso, si esto último sucede, el llenado realizado por la maquina es incorrecto, y deberá dársele mantenimiento preventivo o correctivo.

La mejor máquina, será aquella que vacié de manera constante nueve kilogramos en las bolsas, es decir la mejor máquina es aquella que no tenga variación (o que su variación sea menor).

Valorando el argumento del párrafo anterior, se requiere una medida estadística que mida la variación o dispersión de los datos. A continuación se presentan algunas de dichas medidas de dispersión.

Tabla 1.4 Algunas medidas de dispersión

Varianza Poblacional $Var(x)=\frac{\sum_{i=1}^{n}(x_i-M_A)^2}{n}$	**Varianza Muestral** $Var(x)_m=\frac{\sum_{i=1}^{n}(x_i-M_A)^2}{n-1}$	**Desviación estándar Poblacional** $DE=\sqrt{\frac{\sum_{i=1}^{n}(x_i-M_A)^2}{n}}$
Desviación estándar Muestral $DE_m=\sqrt{\frac{\sum_{i=1}^{n}(x_i-M_A)^2}{n-1}}$	**Desviación absoluta** $DA=\frac{\sum_{i=1}^{n}\lvert x_i-M_A\rvert}{n}$	**Rango de la variable** $R=V_M-V_m$ $V_M=Variante\ mayor$ $V_m=Variante\ menor$

Problema 3.

Considere el conjunto de datos 4, 3, 6, 6, 9, determine los valores para la varianza poblacional, la varianza muestral, la deviación estándar poblacional, la desviación estándar muestral, la desviación absoluta.

De los datos, se tiene que la media aritmética tiene un valor de 5.6

Varianza Poblacional

$$Var(x)=\frac{\sum_{i=1}^{n}(x_i-M_A)^2}{n}=$$

$$=\frac{(4-5.6)^2+(3-5.6)^2+(6-5.6)^2+(6-5.6)^2+(9-5.6)^2}{5}=$$

$$=\frac{(-1.6)^2+(-2.6)^2+(0.4)^2+(0.4)^2+(3.4)^2}{5}=\frac{21.2}{5}=4.24$$

Varianza Muestral

$$Var(x)_m=\frac{\sum_{i=1}^{n}(x_i-M_A)^2}{n-1}\frac{21.2}{5-1}=5.3$$

Desviación estándar Poblacional

$$DE=\sqrt{\frac{\sum_{i=1}^{n}(x_i-M_A)^2}{n}}=\sqrt{4.24}\approx 2.0591$$

Desviación estándar muestral

$$DE_m=\sqrt{\frac{\sum_{i=1}^{n}(x_i-M_A)^2}{n-1}}\approx\sqrt{5.3}=2.3022$$

Desviación absoluta

$$DA=\frac{\sum_{i=1}^{n}\left|x_i-M_A\right|}{n}=$$

$$\frac{|4-5.6|+|3-5.6|+|6-5.6|+|6-5.6|+|9-5.6|}{5}=$$

$$\frac{1.6+2.6+0.4+0.4+3.4}{5}=\frac{8.4}{5}=1.68$$

ACTIVIDAD 7

Determine los valores para la varianza poblacional, la varianza muestral, la deviación estándar poblacional, la desviación estándar muestral, la desviación absoluta de los datos: 2, 3, 3, 5, 5, 6. La media aritmética tiene un valor de 4 unidades.

Solución.

varianza poblacional	$Var(x)=\dfrac{\sum_{i=1}^{n}(x_i-M_A)^2}{n}=$ $=\dfrac{(2-4)^2+(3-4)^2+(3-4)^2+(5-4)^2+(5-4)^2+(6-4)^2}{6}=$ $=\dfrac{(-2)^2+(-1)^2+(1)^2+(1)^2+(1)^2+(2)^2}{6}=\dfrac{12}{6}=2$
varianza mues	$Var(x)=\dfrac{\sum_{i=1}^{n}(x_i-M_A)^2}{n-1}=$ $=\dfrac{(2-4)^2+(3-4)^2+(3-4)^2+(5-4)^2+(5-4)^2+(6-4)^2}{6-1}=$ $=\dfrac{(-2)^2+(-1)^2+(1)^2+(1)^2+(1)^2+(2)^2}{5}=\dfrac{12}{5}=2.4$
deviación estándar poblacional	$DE=\sqrt{\dfrac{\sum_{i=1}^{N_I}(x_i-M_A)^2}{n}}=\sqrt{2}\approx 1.4142$
desviación estándar mues	$DE_m=\sqrt{\dfrac{\sum_{i=1}^{N_I}(x_i-M_A)^2}{n-1}}\approx\sqrt{2.4}=1.5492$
la desvia absoluta	$DA=\dfrac{\sum_{i=1}^{N_I}\lvert x_i-M_A\rvert}{n}=$ $\dfrac{\lvert 2-4\rvert+\lvert 3-4\rvert+\lvert 3-4\rvert+\lvert 5-4\rvert+\lvert 5-4\rvert+\lvert 6-4\rvert}{6}=$ $\dfrac{2+1+1+1+1+2}{6}=\dfrac{8}{6}\approx 1.3334$

La utilidad del coeficiente de variación estriba en que se pueden hacer comparaciones entre diferentes poblaciones de datos, de esta forma poder determinar si una población crece más rápido que otra o bien en cuestiones económicas donde es más conveniente realizar inversiones financieras.

ACTIVIDAD 8

Calcula la varianza poblacional y la varianza muestral de las calificaciones correspondientes a los estudiantes A, B, C, D, considera la tabla de la ***Tabla 1.3 calificaciones de alumnos***, y muestre que solo un estudiante es el mejor.

Solución.

Estudiante A	$Var(x)=\frac{\sum_{i=1}^{n}(x_i-M_A)^2}{n}=\frac{(8-9)^2+(9-9)^2+(10-9)^2}{3}=$ $=\frac{(-1)^2+(0)^2+(1)^2}{3}=\frac{2}{3}$
Estudiante B	$Var(x)=\frac{\sum_{i=1}^{n}(x_i-M_A)^2}{n}=\frac{(10-9)^2+(9-9)^2+(8-9)^2}{3}=$ $=\frac{(1)^2+(0)^2+(-1)^2}{3}=\frac{2}{3}$
Estudiante C	$Var(x)=\frac{\sum_{i=1}^{n}(x_i-M_A)^2}{n}=\frac{(9-9)^2+(9-9)^2+(9-9)^2}{3}=$ $=\frac{(0)^2+(0)^2+(0)^2}{3}=\frac{0}{3}=0$
Estudiante D	$Var(x)=\frac{\sum_{i=1}^{n}(x_i-M_A)^2}{n}=\frac{(10-9)^2+(8-9)^2+(9-9)^2}{3}=$ $=\frac{(1)^2+(-1)^2+(0)^2}{3}=\frac{2}{3}$
	El mejor estudiante es "C"

El mejor estudiante es el que tiene menor varianza (desde el punto de vista estadístico y desde el punto de vista de la de la corriente de aprendizaje, tecnología educativa.

La varianza no puede ser menor a cero si se observa la expresión analítica (formula), el valor mínimo posible es el cero, más adelante al revisar temas de estadística posteriores a este nivel, se buscaran fórmulas para determinar la media y desviación estándar (estimadores) con las propiedad de que proporcionen la menor varianza al cual se le llamara el más eficiente.

Por otra parte existe un método abreviado para calcular la varianza, se conoce con el nombre de formula reducida de la varianza. Se procede a deducirla:

La definición operativa de varianza es:	$Var(x)=\frac{\sum_{i=1}^{n}(x_i-M_A)^2}{n}$
Desarrollando el binomio cuadrado	$Var(x)=\frac{\sum_{i=1}^{n}\left[x_i^2-2M_Ax_i+M^2{}_A\right]}{n}$
Reescribiendo:	$Var(x)=\frac{1}{n}\sum_{i=1}^{n}\left[x_i^2-2M_Ax_i+M^2{}_A\right]$
Aplicando la suma a cada term	$Var(x)=\frac{1}{n}\left[\sum_{i=1}^{n}x_i^2+\sum_{i=1}^{n}\left[-2M_Ax\right]+\sum_{i=1}^{n}M^2{}_A\right]$
Aplicando ley distributiva para $\frac{1}{n}$	$Var(x)=\frac{1}{n}\sum_{i=1}^{n}x_i^2+\frac{1}{n}\sum_{i=1}^{n}\left[-2M_Ax\right]+\frac{1}{n}\sum_{i=1}^{n}M^2{}_A$
Aplicando propiedades de la suma:	$Var(x)=\frac{1}{n}\cdot\sum_{i=1}^{n}x_i^2-2M_A\frac{1}{n}\cdot\sum_{i=1}^{n}x_i+\frac{1}{n}\cdot n\cdot M^2{}_A$
el termino $\frac{1}{n}\cdot\sum_{i=1}^{n}x_i$	Es precisamente la media aritmética
simplificando	$Var(x)=\frac{1}{n}\cdot\sum_{i=1}^{n}x_i^2-2M^2{}_A+M^2{}_A$

Reduciendo, tenemos la llamad formula reducida de la varianza.	$Var(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^{\,2} - M^2{}_A$

ACTIVIDAD 9

Considere los datos cuyos valores son: 2, 3, 3, 5, 5, 6, presentados en la ACTIVIDAD 7, calcule la varianza con la formula reducida y muestre que el valor es igual a dos, es decir $Var(x) = 2$. El valor de media aritmética es de cuatro unidades.

Solución.

Sustituyendo datos en:	$Var(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^{\,2} - M^2{}_A$
Tenemos:	$Var(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^{\,2} - M^2{}_A = \frac{1}{6} \cdot \left[2^2 + 3^2 + 3^2 + 5^2 + 5^2 + 6^2\right] - (4)^2$
Simplificando:	$Var(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^{\,2} - M^2{}_A = \frac{108}{6} - 16 = 18 - 16 = 2$
	$Var(x) = 2$

ACTIVIDAD 10

Determine el valor de la varianza y de la desviación estándar, para los datos:{1, 2, 3, 4, 5, 6, 6, 7, 8, 9}, usando el método de la formula reducida.

Solución.

Tabla 1.6 cálculo de la formula reducida de la varianza

	x_i	x_i^2
	1	**1**
	2	**4**
	3	**9**
	4	**16**
	5	**25**
	6	**36**
	6	**36**
	7	**49**
	8	**64**
	9	**81**
suma	**51**	**321**
	$M_A = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{51}{10} = 5.1$	$\frac{1}{n}\sum_{i=1}^{n} x^2{}_i = \frac{321}{10} = 32.1$

La fórmula reducida de la varianza:

$$Var(x) = \frac{1}{n}\cdot\sum_{i=1}^{n} x_i^{\,2} - M^2{}_A =$$

$$= \; 32.1 - (5.1)^2 = 6.09$$

La Desviación estándar correspondiente es:

$$DE = \sqrt{Var(x)} = \sqrt{6.09} = 2.4678$$

Se debe tener cuidado con el uso de la formula reducida y no confundir la suma de los cuadrados de los datos con la suma al cuadrado de los datos.

1.2.1 ALGUNAS MEDIDAS DE DISPERSIÓN RELATIVA

Cuando existe la necesidad e comparar la variación de datos, pero estos pertenecen a poblaciones distintas, se requiere una medida de la variabilidad que haga caso omiso de las unidades físicas. Por ejemplo un inversionista desea saber si es mejor invertir su dinero en acciones de renta variable que cotizan en la bolsa de valores, en dólares, euros o comprar oro amonedado, o cuando un biólogo necesita saber qué población animal o vegetal crece más rápido; por ejemplo determinada especie de peces, insectos, ratones, cultivo de maíz mundial.

Para determinar si los ratones o insectos se pudieran considerar como una plaga, es necesario disponer de una herramienta que caractericen esa variación, independientemente de los tipos de objetos o de las unidades de medición.

Por otra parte el coeficiente de variación se utiliza para medir la homogeneidad de la distribución de datos, a mayor valor existe mayor dispersión (error) o incertidumbre, a menor valor, se considera que la distribución de los datos es homogénea.

Tabla 1.7 fórmulas de los coeficiente de variación y variación porcentual.

Coeficiente de variación	$CV = \frac{DE}{M_A}$
Coeficiente de variación Porcentual	$CV\% = \frac{DE}{M_A} \times 100\%$

ACTIVIDAD 11

Calcular el coeficiente de variación y el coeficiente de variación porcentual para los siguientes datos: 2, 3, 3, 5, 5, 6. Por simplicidad considérelos como una población.

Solución.

Para calcular el coeficiente de variación y el coeficiente de variación Porcentual, necesitamos la media aritmética y la desviación estándar.

Primero se determina la media, la varianza y la desviación estándar:

$$M_A = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2+3+3+5+5+6}{6} = \frac{24}{6} = 4$$
$$Var(x) = \frac{\sum_{i=1}^{N_I} (x_i - M_A)^2}{n} =$$ $$= \frac{(2-4)^2+(3-4)^2+(3-4)^2+(5-4)^2+(5-4)^2+(6-4)^2}{6} =$$ $$= \frac{(-2)^2+(-1)^2+(-1)^2+(1)^2+(1)^2+(2)^2}{6} =$$ $$= \frac{4+1+1+1+1+4}{6} == 2$$
$$DE = \sqrt{Var(x)} = \sqrt{2} = 1.4142$$

Coeficiente de variación	$$CV = \frac{DE}{M_A} = \frac{1.4142}{4} \approx 0.3536$$
Coeficiente de variación Porcentual	$$CV\% = \frac{DE}{M_A} \times 100\% \quad , \quad CV\% = 0.3536 \times 100\% = 35.36\%$$

ACTIVIDAD 12

La información financiera dados por Banco de México acerca de las variaciones del dólar vs peso, euro contra el peso, yen japonés vs peso y centenario en el mes de Septiembre del año 2012, son las siguientes:

Tabla 1.7 precio de divisas

	Media aritmética	varianza
Dólar	12.83	0.90
Euro	16.69	2.25
Yen japonés	0.1642	0.0043
Centenario	26,000	4000000

Calcular el coeficiente de variación y el coeficiente de variación porcentual, para cada una de los instrumentos financieros y determinar cuál es el de mejor variación.

Solución.

	Desviación estándar	**Media aritmética**	**Coeficiente de variación** $CV = \frac{DE}{M_A}$	**Coeficiente de Variación Porcentual** $CV\% = \frac{DE}{M_A} \times 100\%$
Dólar	0.9487	12.83	$CV = \frac{0.9487}{12.83} = 0.0739$	7.39 %
Euro	1.5	16.69	$CV = \frac{1.5}{16.69} = 0.0899$	8.99 %
Yen japonés	0.0655	0.1642	$CV = \frac{0.0655}{0.1642} = 0.3989$	39.89 %

Centenario	2000	26,000	$CV = \frac{2000}{26000} = 0.0769$	7.69 % ***Tiene menor*** ***Variación***

ACTIVIDAD 13 Refuerzo al aprendizaje de medidas centrales, dispersión absoluta y relativa.

La información que a continuación se indica, representa el número de refrescos consumidos por semana, después de realizar una encuesta a un grupo de 6 estudiantes.

Tabla 1.8 número de refrescos consumidos por semana

estudiante 1	estudiante 2	estudiante 3	estudiante 4	estudiante 5	estudiante 6
2	**3**	**6**	**4**	**5**	**5**

Relacione la columna que tienen paréntesis con las de los números, de forma que la afirmación sea la correcta.

( ) valor de la Media aritmética	1. 6.000	11. 1.4142
( ) valor de la mediana	2. 4.5000	12. 3.7113
( ) valor de la moda	3. 1.5678	13. 1.3436
( ) valor de la Media armónica	4. 4.000	14. 4.0000
( ) valor de la Media geométrica	5. 1.8053	15. 5.4000
( ) valor de la Varianza Poblacional	6. 1.1667	16. 3.9149
( ) valor de la desviación estándar	7. 0.3225	17. 0.2543
( ) valor de la Desviación estándar muestral.	8. 1.4719	18. 1.6543

( )	valor del rango	9.	3.1009	19.	4.1667
( )	valor del coeficiente de variación	10.	5.0000	20.	0.2000
( )	valor de la Desviación absoluta				

Solución.

(19)	valor de la Media aritmética	1.	6.000	11.	1.4142
(2)	valor de la mediana	2.	4.5000	12.	3.7113
(10)	valor de la moda	3.	1.5678	13.	1.3436
(12)	valor de la Media armónica	4.	4.000	14.	4.0000
(16)	valor de la Media geométrica	5.	1.8053	15.	5.4000
(5)	valor de la Varianza Poblacional	6.	1.1667	16.	3.9149
(13)	valor de la desviación estándar	7.	0.3225	17.	0.2543
(8)	valor de la Desviación estándar muestral.	8.	1.4719	18.	1.6543
(14)	valor del rango	9.	3.1009	19.	4.1667
(7)	valor del coeficiente de variación	10.	5.0000	20.	0.2000
(6)	valor de la Desviación absoluta				

Media aritmética	$M_A = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{2+3+6+4+5+5}{6} = \frac{25}{6} = 4.1667$
mediana	acomodando datos en forma creciente: 2,3,4,5,5,6 por ser par el número de datos, seleccionamos los datos que ocupan el ordinal tercero y cuarto ordinal son; 4 y 5. $M_E = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{4+5}{2}$. La mediana de las calificaciones es $M_E = 4.5$
moda	El valor con mayor frecuencia es el 5, se repite dos veces entonces: $M_O = 5$
media armónica	$M_H = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} \ldots + \frac{1}{x_n}} =$ $= \frac{6}{\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{5}+\frac{1}{6}} \approx 3.7113$
media geométrica	$M_G = \sqrt[n]{\pi_{i=1}^{n} x_i} = \sqrt[n]{x_1 \cdot x_2 \ldots x_n} =$ $= \sqrt[6]{3 \cdot 2 \cdot 4 \cdot 5 \cdot 5 \cdot 6} \approx 3.9148$
La fórmula reducida de la variar	$Var(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 - M^2{}_A =$ $= \frac{1}{6}\left\{2^2 + 3^2 + 4^2 + 5^2 + 5^2 + 6^2\right\} - (4.1667)^2 =$ $= \frac{115}{6} - 17.3613 \approx 1.8053$

La Desviación estánd correspondiente es:	$DE=\sqrt{Var(x)}=\sqrt{1.8053}=1.3436$
La Desviación estánd muestral	$Var(x)_m=\frac{n}{n-1}\mathrm{var}(x)=\frac{6}{6-1}(1.8053)$ $=2.1664$ $DE_m=\sqrt{Var(x)_m}=\sqrt{2.1664}\approx 1.4719$
valor del rango	$R=V_M-V_m=6-2=4$ $V_M=Variante\ mayor$ $V_m=Variante\ menor$
valor del coeficiente de variación	$CV=\frac{DE}{M_A}=\frac{1.3436}{4.1667}\approx 0.3225$

valor de la Desviación absoluta

$$DA=\frac{\sum_{i=1}^{n}\left|x_i-M_A\right|}{n}=$$

$$\frac{|2-4.1667|+|3-4.1667|+|4-4.1667|+|5-4.1667|+|5-4.1667|+|6-4.1667|}{6}=$$

$$=\frac{7}{6}\approx 1.1667$$

ACTIVIDAD 14 Refuerzo al aprendizaje de medidas centrales, dispersión absoluta y relativa.

Se pesan 40 bolsitas de papas "Sabritas", el contenido en gramos de la etiqueta indica que el producto envasado debe pesar 37 gramos, la información disponible es: {34, 35, 35, 36, 36, 36, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 38, 38, 38, 38, 38, 38, 38, 38, 38, 38}

Relacione la columna que tienen paréntesis con las de los números, de forma que la afirmación sea la correcta.

( )	valor de la Media aritmética	1.	34.0000	11.	0.7500
( )	valor de la Media mediana	2.	36.97	12.	39.0000
( )	valor de la Media moda	3.	36.9790	13.	33.0000
( )	valor de la Media armónica	4.	4.0000	14.	0.8660
( )	valor de la Media geométrica	5.	37.0000	15.	0.8077
( )	valor de la Varianza Poblacional	6.	0.9486	16.	36.9896
( )	valor de la desviación estándar poblacional	7.	1.0987	17.	0.8321
( )	valor de la Desviación estándar muestral.	8.	32.9876	18.	35.9876
( )	valor del rango	9.	35.0098	19.	5.9893
( )	valor del coeficiente de variación	10.	0.0324	20.	9.4500
( )	valor de Desviación media absol			21	0.0423

Solución.

(5)	valor de la Media aritmética	1.	34.0000	11.	0.7500
(5)	valor de la Media mediana	2.	36.97	12.	39.0000
(5)	valor de la Media moda	3.	36.9790	13.	33.0000
(2)	valor de la Media armónica	4.	4.0000	14.	0.8660
(16)	valor de la Media geométrica	5.	37.0000	15.	0.8077
(11)	valor de la Varianza Poblacional	6.	0.9486	16.	36.9896
(14)	valor de la desviación estándar poblacional	7.	1.0987	17.	0.8321
(6)	valor de la Desviación estándar muestral.	8.	32.9876	18.	35.9876
(4)	valor del rango	9.	35.0098	19.	5.9893
(21)	valor del coeficiente de variación	10.	0.0324	20.	9.4500
(20)	valor de Desviación media absol			21	0.0423

Media Aritmética	$M_A = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{34+35+...+38}{40} = \frac{34+2(35)+3(36)+24(37)+10(38)}{40} = \frac{1480}{40} = 37$
	Los datos están acomodando datos en forma creciente: por ser par el número de datos, seleccionamos los datos que ocupan el

mediana	ordinal veinte y el veintiuno ordinal son; 37 y 37. $M_E=\dfrac{x_{\frac{n}{2}}+x_{\frac{n}{2}+1}}{2}=\dfrac{37+37}{2}$. La mediana de los pesos es $M_E=37$
moda	El valor con mayor frecuencia es el 37,: $M_O=37$
media armónica	$M_H=\dfrac{n}{\sum_{i=1}^{n}\frac{1}{x_i}}=\dfrac{n}{\frac{1}{x_1}+\frac{1}{x_2}\ldots+\frac{1}{x_n}}=$ $=\dfrac{40}{\frac{1}{34}+2\cdot\frac{1}{35}+3\cdot\frac{1}{36}+24\cdot\frac{1}{37}+10\cdot\frac{1}{38}}\approx 36.97$
media geométrica	$M_G=\sqrt[n]{\pi_{i=1}^{n} x_i}=\sqrt[n]{x_1\cdot x_2\ldots x_n}=$ $=\sqrt[40]{34\cdot 35^2\cdot 36^3\cdot 37^{24}\cdot 38^{10}}\approx 36.9896$
La fórmula reducida de la varianza, para calcular la varianza poblacional	$Var(x)=\dfrac{1}{n}\cdot\sum_{i=1}^{n}x_i^2-M^2{}_A=$ $=\dfrac{1}{40}\left\{34^2+2\cdot(35)^2+3\cdot(36)^2+24\cdot(37)^2+10\cdot(38)^2\right\}-(37)^2=$ $=\dfrac{54790}{40}-1369\approx 0.75$
La Desviación estándar correspondiente es:	$DE=\sqrt{Var(x)}=\sqrt{0.75}=0.8660$
La Desviación estándar muestral es:	$Var(x)_m=\dfrac{n}{n-1}Var(x)=\dfrac{6}{6-1}(0.75)$ $=0.9$ $DE_m=\sqrt{Var(x)_m}=\sqrt{0.9}\approx 0.9486$

<table>
<tr>
<td>valor del rango</td>
<td>$$R = V_M - V_m = 38 - 34 = 4$$
$$V_M = Variante\ mayor$$
$$V_m = Variante\ menor$$</td>
</tr>
<tr>
<td>valor del
coeficiente de
variación</td>
<td>$$CV = \frac{DE}{M_A} = \frac{0.8660}{37} \approx 0.0423$$</td>
</tr>
<tr>
<td colspan="2">valor de la Desviación absoluta
$$DA = \frac{\sum_{i=1}^{N_I} \left| x_i - M_A \right|}{n} =$$
$$\frac{\left|34-37\right| + 2\cdot\left|35-37\right| + 3\cdot\left|36-37\right| + 24\cdot\left|37-37\right| + 10\cdot\left|38-37\right|}{40} =$$
$$= \frac{3 + 2\cdot(2) + 3\cdot(1) + 24\cdot(0) + 10\cdot(38)}{40} = \frac{390}{40} \approx 9.45$$</td>
</tr>
</table>

1.3 LAS MEDIDAS DE POSICIÓN PARA DATOS NO AGRUPADOS. LOS CUANTILES; CUARTILES, DECILES Y PORCENTILES.

El término "cuantil" fue usado por primera vez por Maurice George Kendall en 1940. El cuantil marca un corte de modo que una proporción p de valores de la población es menor o igual que la proporción de dicho cuantil.

Los cuantiles se usan para dividir una población de datos en partes iguales. Los más comunes son:

a. Los Cuartiles; Q_1, Q_2, Q_3 que dividen a la distribución en cuatro partes, corresponden a la división de la población de datos en el 25%, 50% y 75%.

b. Los **Deciles**, $D_1, D_2, D_3 ..., D_9$ que dividen a la distribución en diez partes; es decir en el 10%, 20%,…90%

c. Los **Percentiles**, $P_1, ..., P_{99}$ que dividen a la distribución en cien partes, es decir; 1%, 2%,…,99%

El cálculo de estos valores, siempre será con una aproximación, en ocasiones muy baja, y habrá que esperar hasta la construcción de la tabla de datos agrupados, donde será posible determinarlos con mayor precisión, por lo pronto nos conformamos con que estas partes sean aproximadamente iguales.

Se sabe que existen nueve métodos diferentes, que conducen a resultados diferentes, pero muy aproximados entre sí para determinarlos.

1.3.1 EL RANGO Y LOS CUANTILES

En ocasiones es necesario dividir la información en partes para su estudio, por ejemplo los cuartiles dividen la información en cuatro partes iguales. Los Deciles dividen la información de los datos en diez partes iguales. Los porcentiles dividen la información en cien partes iguales, de manera general estas divisiones se denominan cuantiles.

Problema 4.

Considere el conjunto de datos 0, 1, 1, 1, 1, 1, 1, 2 , 2 , 2 , 2 , 2 , 2 , 2, 3 , 3, 4, 3, 6, 6, 9, determine los valores del rango, los cuartiles; Q_1, Q_2, Q_3 ,el cuartil 2 corresponde al valor de la mediana.

El Rango (también llamado longitud o recorrido de la variable) se puede calcular como la diferencia del dato de mayor valor, menos el dato de menor valor:

$R = V_M - V_m$, por lo que $R = 6 - 0 = 6$.

Para calcular el cuartil número uno, dividimos el número de datos entre 4 (o multiplicamos por 0.25, tenemos 21 datos, así que $21 \times 0.25 = 5.25$, como la definición dice que debemos tomar el número ordinal menor o igual, le corresponde al dato en el ordinal 5, que en este caso es el número

$Q_1 = 1$.

Para calcular el cuartil número dos, dividimos el número de datos entre 2 (o multiplicamos por 0.5, tenemos 21 datos, así que $21 \times 0.5 = 10.5$, como la definición dice que debemos tomar el número ordinal menor o igual, le corresponde al dato en el ordinal 10, que en este caso es el número

$Q_2 = 2$.

Para calcular el cuartil número tres, multiplicamos el número de datos por 3/4 (o multiplicamos por 0.75, tenemos 21 datos, así que $21 \times 0.75 = 15.75$, como la definición dice que debemos tomar el número ordinal menor o igual, le corresponde al dato en el ordinal 15, que en este caso es el número

$Q_3 = 3$.

Aunque los índices coinciden con los números, solo es una afortunada coincidencia.

ACTIVIDAD 15

Considere el conjunto de datos 0, 1, 1, 1, 1, 1, 1, 2 , 2 , 2 , 2 , 2 , 2 , 2, 3 , 3, 4, 3, 6, 6, 9, determine los valores de los deciles, D_1, D_7, D_6, D_9.

Solución.

Siguiendo con el mismo proceso indicado para la obtención de los cuartiles, en este caso dividiremos la información entre diez. Se tienen 21 datos ordenados de menor a mayor.

$21 \times 0.10 = 2.1$	corresponde el ordinal 3	$D_1 = 1$
$21 \times 0.60 = 12.6$	corresponde el ordinal 13	$D_6 = 1$
$21 \times 0.70 = 14.7$	corresponde el ordinal 15	$D_7 = 1$
$21 \times 0.90 = 18.9$	corresponde el ordinal 19	$D_9 = 6$

ACTIVIDAD 16

Considere el conjunto de datos 0, 1, 1, 1, 1, 1, 1, 2 , 2 , 2 , 2 , 2 , 2 , 2, 3 , 3, 4, 3, 6, 6, 9, determine los valores de los porcentiles, $P_{13}, P_{22}, P_{67}, P_{97}$

Solución.

Siguiendo con el mismo proceso indicado para la obtención de los cuartiles, en este caso dividiremos la información entre diez. Se tienen 21 datos ordenados de menor a ma

$P_{13} = 21 \times 0.13 = 2.73$	corresponde el ordinal 3	$P_{13} = 1$
$P_{22} = 21 \times 0.60 = 4.62$	corresponde el ordinal 5	$P_{22} = 1$
$P_{67} = 21 \times 0.67 = 14.07$	corresponde el ordinal 15	$P_{67} = 1$

$P_{97} = 21 \times 0.97 = 20,37$	corresponde el ordinal 21	$P_{97} = 6$

ACTIVIDAD 17

Explique cómo se construye un diagrama de caja y bigote, considere los datos 0, 1, 1,

1, 1, 1, 1, 2 , 2 , 2 , 2 , 2 , 2 , 2, 3 , 3, 4, 3, 6, 6, 9.

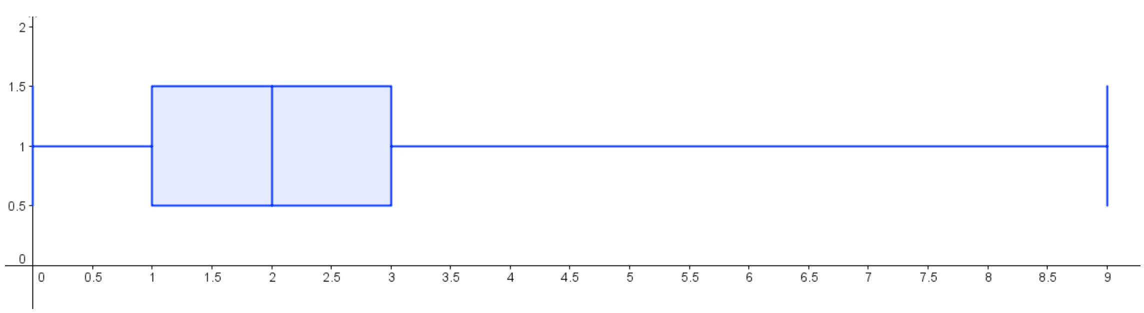
Solución.

1. Se determina el valor mínimo y máximo del conjunto de datos, en este caso es; uno y el nueve.

2. se calculan los cuartiles Q_1, Q_2, Q_3 ,los cuales sirven para construir la caja .
3. los bigotes son segmentos de recta del valor mínimo a Q_1 , y de Q_3 al valor Máximo.

ACTIVIDAD 18

Considere la gráfica de caja y bigote, mostrada abajo, determine los cuartiles y el rango de los datos y diga si la distribución de los datos es simétrica.



Solución.

1. el valor mínimo y máximo del conjunto de datos, en este caso es; uno y el nueve.

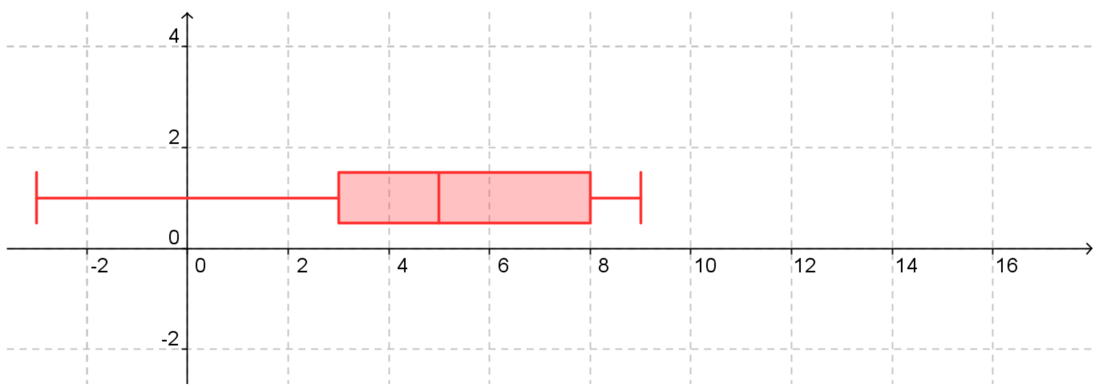
2. los cuartiles son: $Q_1 = 1, \quad Q_2 = 2, \quad Q_3 = 3$

3. el bigote izquierdo tiene longitud uno , el de lado derecho es de seis

4. la distribución de los datos <u>es simétrica</u>. la diferencia entre el primer cuartil y el segundo cuartil ,debe ser la misma que la diferencia entre el segundo cuartil y el tercer cuartil, recuerde que la mediana es igual al 2o cuartil.

ACTIVIDAD 19

Considere la figura. Determine el Rango de los datos, los valores de los cuartiles uno, dos y tres. ¿Se puede considerar simétrica la distribución de datos?



Solución.

1. El valor mínimo y máximo del conjunto de datos, en este caso es; -3 y el 9. El rango de los datos : $R = V_M - V_m = 9 - (-3) = 12$

2. Los cuartiles son: $Q_1 = 3, \quad Q_2 = 5, \quad Q_3 = 8$

3. No, la distribución de los datos <u>no es simétrica</u>. la diferencia entre el primer cuartil y el segundo cuartil ,debe ser la misma que la diferencia entre el segundo cuartil y el tercer cuartil, recuerde que la mediana es igual al 2o cuartil.

ACTIVIDAD 20

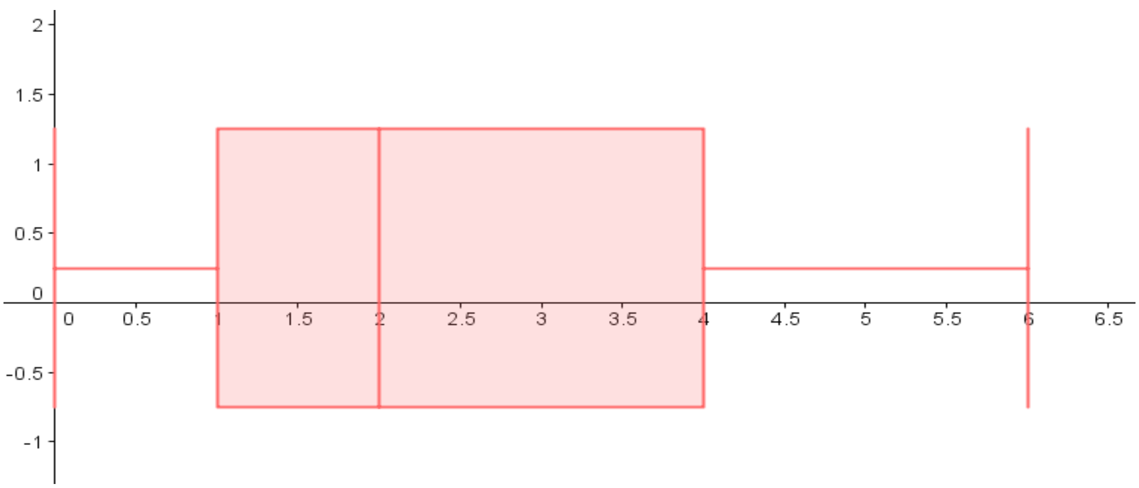
Considere la gráfica de caja y bigote, mostrada abajo, determine.

a. El valor máximo y mínimo del conjunto de datos.

b. El rango.

c. Los cuartiles Q_1, Q_2, Q_3

d. Determine si la distribución de los datos es simétrica o diga si es sesgada a la derecha o a la izquierda. Argumente su respuesta.



Solución.

a. El valor máximo es 6 y mínimo es 0.

b. El rango. $R = V_M - V_m = 6 - 0 = 6$

c. Los cuartiles $Q_1 = 1, \quad Q_2 = 2, \quad Q_3 = 4$

d. No, la distribución de los datos no es simétrica.
la diferencia entre el primer cuartil y el segundo cuartil ,debe ser la misma

que la diferencia entre el segundo cuartil y el tercer cuartil, recuerde usted que la mediana es precisamente al 2o cuartil.

ACTIVIDAD 21

Investigue como se construye un diagrama de tallo y hoja y aplique ese resultado para los siguientes datos, utilice el esqueleto de tabla que se muestra abajo:

Tabla 1.9 datos para la construcción del diagrama de tallo y hoja.

1.1	1.4.	2.8	4.1	4.6	5.1
1.3.	2.5	2.9	4.2	4.6	5.2
1.4	2.6	3.6	4.3	4.7	5.3
1.4	2.6	3.6	4.4	5.1	5.4
1.4	2.7	3.7	4.5	5.1	5.5
1.4	2.7	3.7	4.5	5.1	5.5

Utilice este espacio para construir el diagrama.

Solución.

1	0.1	0.3	0.4	0.4	0.4	0.4	0.4		
2	0.5	0.6	0.6	0.7	0.7	0.8	0.9		
3	0.6	0.6	0.7	0.7					
4	0.1	0.2	0.3	0.4	0.5	0.5	0.6	0.6	0.7
5	0.1	0.1	0.1	0.1	0.2	0.3	0.4	0.5	0.5

ACTIVIDAD 22

Construya un diagrama de tallo y hoja y aplique ese resultado para los siguientes datos, utilice el esqueleto de tabla que se muestra abajo:

1.1	1.4.	2.8	4.1	5.1	6.1
1.3.	2.5	2.9	4.2	5.2	6.2
1.3.	2.5	2.9	4.2	5.2	6.2
1.4	2.6	3.6	4.3	5.3	7.3
1.4	2.6	3.6	4.3	5.3	7.3

1.4	2.6	3.6	4.4	5.4	7.4
1.4	2.7	3.7	4.5	5.5	7.5
1.4	2.7	3.7	4.5	6.1	7.5

Solución.

1	0.1	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.4
2	0.5	0.5	0.6	0.6	0.7	0.7	0.8	0.9	0.9
3	0.6	0.6	0.6	0.7	0.7				
4	0.1	0.2	0.2	0.3	0.3	0.4	0.5	0.5	
5	0.1	0.2	0.2	0.3	0.3	0.4	0.5		
6	0.1	0.1	0.2	0.2					
7	0.3	0.3	0.4	0.5	0.5				

ACTIVIDAD 23

La información que a continuación se indica, representa el número de refrescos consumidos por semana. Después de realizar una encuesta a un grupo de 20 estudiantes, se obtuvo la siguiente información:{2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6}. Relacione ambas columnas de manera correcta.

( )	valor del cuartil 1		1.	5.00
( )	valor del cuartil 2		2.	2.00
( )	valor del cuartil 3		3.	2.50
( )	valor del cuartil 4		4.	5.50
( )	Rango		5.	4.00
( )	Decil 6		6.	3.00
( )	Decil 8		7.	4.10

( )	Porcencil 46		9.	6.54
( )	Porcencil 89		10.	2.54
( )	Porcencil 17		11.	6.00
			12.	5.05
			13.	5.69
			14.	4.12

Solución.

(2)	valor del cuartil 1		1.	5.00
(6)	valor del cuartil 2		2.	2.00
(5)	valor del cuartil 3		3.	2.50
(11)	valor del cuartil 4		4.	5.50
(5)	Rango		5.	4.00
(5)	Decil 6		6.	3.00
(1)	Decil 8		7.	4.10
(6)	Porcencil 46		9.	6.54

(1)	Porcencil 89		10.	2.54
(2)	Porcencil 17		11.	6.00
			12.	5.05
			13.	5.69
			14.	4.12

valor del cuartil 1	Multiplicamos por 4 el número de datos, tenemos 20 datos, así que $20\times0.25=5$, debemos tomar el número ordinal menor o igual al quinto dato , es decir: $Q_1=2$
valor del cuartil 2	Multiplicamos por 4 el número de datos, tenemos 20 datos, así que $20\times0.50=10$, debemos tomar el número ordinal menor o igual al décimo dato, $Q_2=3$
valor del cuartil 3	Multiplicamos por 4 el número de datos, tenemos 20 datos, así que $20\times0.75=15$, debemos tomar el número ordinal menor o igual al quinceavo dato, $Q_3=4$
valor del cuartil 4	Multiplicamos por 4 el número de datos, tenemos 20 datos, así que $20\times1.00=20$, debemos tomar el número ordinal menor o igual al vigésimo dato, $Q_4=6$
Rango	El Rango (también llamado longitud o recorrido de la variable) se puede calcular como la diferencia del dato de mayor valor, menos el dato de menor valor: $R=V_M-V_m$, por lo que $R=6-2=4$.
Decil 6	Multiplicamos por 0.6 el número de datos, tenemos 20 datos, así que $20\times0.60=12$, debemos tomar el número ordinal menor o igual al doceavo dato , es decir: $D_6=4$
Decil 8	Multiplicamos por 0.8 el número de datos, tenemos 20 datos, así que $20\times0.80=16$, debemos tomar el número ordinal menor o igual al dieciseisavo dato , es decir: $D_8=5$
Porcencil 46	Multiplicamos por 0.46 el número de datos, tenemos 20 datos,

	así que $20 \times 0.46 = 9.2 \approx 9$, debemos tomar el número ordinal menor o igual al noveno dato , es decir: $P_{46} = 3$
Porcencil 89	Multiplicamos por 0.89 el número de datos, tenemos 20 datos, así que $20 \times 0.89 = 17.8 \approx 17$, debemos tomar el número ordinal menor o igual al diecisieteavo dato , es decir: $P_{89} = 5$
Porcencil 17	Multiplicamos por 0.17 el número de datos, tenemos 20 datos, así que $20 \times 0.17 = 3.4 \approx 3$, debemos tomar el número ordinal menor o igual al noveno dato , es decir: $P_{17} = 2$

ACTIVIDAD 24

Considere los datos los 50 datos obtenidos como medición del contenido de líquido envasado en botellas cuya etiqueta indica un contenido de 600 mililitros, la información obtenida es:

{586, 586, 586, 586, 586, 586, 589, 589, 589, 589, 589, 589, 589, 589, 590, 590, 590, 590, 590, 590, 590, 590, 593, 595, 596, 597, 597, 598, 599, 599, 599, 599, 599, 599, 599, 599, 599, 599, 599, 599, 600, 600, 600, 600, 600, 600, 600, 600, 600, 600}

Relacione ambas columnas de manera correcta.

( )	valor del cuartil 1	1.	600.0
( )	valor del cuartil 2	2.	589.6
( )	valor del cuartil 3	3.	590.0
( )	valor del cuartil 4	4.	14
( )	Rango	5.	596.0
( )	Decil 3	6.	602.5
( )	Decil 7	7.	596.3
( )	Porcentil 63	9.	589.0

( )	Porcentil 39	10.	598.2
( )	Porcentil 23	11.	589.0
		12.	599.0
		13.	599.9
		14.	599.0

Solución.

(11)	valor del cuartil 1	1.	600.0
(5)	valor del cuartil 2	2.	589.6
(12)	valor del cuartil 3	3.	590.0
(1)	valor del cuartil 4	4.	14
(4)	Rango	5.	596.0
(3)	Decil 3	6.	602.5
(14)	Decil 7	7.	596.3
(14)	Porcentil 63	9.	589.0
(3)	Porcentil 39	10.	598.2
(9)	Porcentil 23	11.	589.0
		12.	599.0
		13.	599.9
		14.	599.0

valor del cuartil 1	Multiplicamos por 0.25 el número de datos, tenemos 50 datos, así que $50\times 0.25 = 12.5 \approx 12$, debemos tomar el número ordinal menor o igual al doceavo dato , es decir: $Q_1 = 589$
valor del cuartil 2	Multiplicamos por 0.50 el número de datos, tenemos 50 datos, así que $50\times 0.50 = 25$, debemos tomar el número ordinal menor o igual al veinticincoavo dato, $Q_2 = 596$
valor del cuartil 3	Multiplicamos por 0.75 el número de datos, tenemos 50 datos, así que $50\times 0.75 = 37.5 \approx 37$, debemos tomar el número ordinal menor o igual al treintaisieteavo dato, $Q_3 = 599$
valor del cuartil 4	Multiplicamos por 1.0 el número de datos, tenemos 50 datos, así que $50\times 1.00 = 50$, debemos tomar el número ordinal menor o igual al cincuentavo dato, $Q_4 = 600$
Rango	El Rango (también llamado longitud o recorrido de la variable) se puede calcular como la diferencia del dato de mayor valor, menos el dato de menor valor: $R = V_M - V_m$, por lo que $R = 600 - 586 = 14$.
Decil 3	Multiplicamos por 0.30 el número de datos, tenemos 50 datos, así que $50\times 0.30 = 15$, debemos tomar el número ordinal menor o igual al quinceavo dato , es decir: $D_3 = 590$
Decil 7	Multiplicamos por 0.70 el número de datos, tenemos 50 datos, así que $50\times 0.70 = 35$, debemos tomar el número ordinal menor o igual al treintaicincoavo dato , es decir: $D_7 = 599$
Porcencil 63	Multiplicamos por 0.63 el número de datos, tenemos 50 datos, así que $50\times 0.63 = 31.5 \approx 31$, debemos tomar el número ordinal menor o igual al treintaiunavo dato , es decir: $P_{63} = 599$
Porcencil 39	Multiplicamos por 0.39 el número de datos, tenemos 50 datos, así que $50\times 0.39 = 19.5 \approx 19$, debemos tomar el número ordinal menor o igual al diecinueveavo dato , es decir: $P_{39} = 590$
Porcencil 23	Multiplicamos por 0.23 el número de datos, tenemos 50 datos, así que $50\times 0.23 = 11.5 \approx 11$, debemos tomar el número ordinal menor o igual al onceavo dato , es decir: $P_{23} = 589$

1.4 ALGUNAS MEDIDAS DE TENDENCIA CENTRAL Y DE DISPERSIÓN PARA DATOS AGRUPADOS.

La agrupación de datos tiene por objetivo determinar las características de la población. Tales como puntos máximos, mínimos, donde es creciente o decreciente, la forma que tiene etc. Según otros autores, se logra el objetivo de resumir los datos.

Dada la tabla de la distribución de frecuencias, es posible hallar medidas de tendencia central, así como también medidas de dispersión.

Algunas de las definiciones necesarias en el proceso de agrupación de datos se enuncian:

La frecuencia es el número de veces en que un dato se repite. f_i

La frecuencia absoluta f_i, es el número de datos que caen en un intervalo o categoría.

La frecuencia relativa $f_{Ri} = \frac{f_i}{n}$, es el número de datos que caen en un intervalo "i"o categ

dividido entre el total de datos.

A continuación se revisaran algunos problemas en donde es necesario, organizar, y agrupar datos en categorías o intervalos

PROBLEMA 5

En la actividad 12 , se dispone de un conjunto de datos, correspondientes a cuarenta bolsitas de papas "Sabritas", el peso en gramos de cada bolsita es: 34, 35, 35, 36, 36, 36, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 37, 38, 38, 38, 38, 38, 38, 38, 38, 38, 38.Organice los datos en una tabla.

Inicialmente se deben ordenar la información de los datos de manera creciente o bien decreciente, una vez hecho esto, para contarlos se debe de ir tachándolos los elementos que caen en determinada categoría, con la finalidad de no contarlos nuevamente y evitar errores. la tabla. En el caso de los datos del problema 5, los datos ya están ordenados y solo se deben contar cuantos caen en determinada categoría. Las categorías están consideradas de antemano, para evitar contratiempos, sin embargo existen algunos algoritmos que permiten seleccionar el número de intervalos más adecuado para la presentación de la información.

Tabla 1.10 datos agrupados; categorías y frecuencia absoluta.

Categoría o Intervalo de Clase Real	Frecuenciaabsoluta
33.5 - 34.5	1
34.5 - 35.5	2
35.5 - 36.5	3
36.5 - 37.5	24
37.5 - 385	10
	La suma de las frecuencias de cada uno de los cinco intervalos, es igual a la totalidad de los datos: $n=\sum_{i=1}^{5} f_i = 1+2+3+24+10=40$

Es necesario obtener las siguientes cantidades; frecuencia absoluta; f_i, Frecuencia relativa; f_{Ri}, frecuencia porcentual; $f_{Ri}\%$, frecuencia acumulada absoluta F_i, frecuencia acumulada relativa; F_{Ri} y la frecuencia acumulada porcentual; $F_{Ri}\%$, las marcas de clase; M_i .

La finalidad de su construcción es la de obtener posteriormente las medidas centrales y medidas de dispersión absoluta y relativa, para los datos agrupados, así como también para la construcción de los gráficos que representaran la información.

Tabla 1.11 elementos para la construcción de la tabla de frecuencias, datos agrupados.

Frecuencia Absoluta: f_i ***"Se obtiene de***	**Frecuencia Relativa:**	**Frecuencia Porcentual :**	**Marca de clase**

contar cuantos datos caen en el inte de clase”	$f_{Ri} = \frac{f_i}{\sum_{I=1}^{N_I} f_i} = \frac{f_i}{n}$	$f_{Ri}\% = \frac{f_i}{\sum_{i=1}^{N_I} f_i} \times 100\% = \frac{f_i}{n} \times 100\%$	$m_i = \frac{L_{S1} + L_{i1}}{2}$
Frecuencia Acumulada Absoluta: F_i $F_i = \sum_{I=1}^{N_I} f_i$	**Frecuencia Acumulada Relativa:** $F_{Ri} = \frac{\sum_{I=1}^{N_I} f_i}{n}$	**Frecuencia Acumulada Porcentual :** $F_i = \frac{\sum_{i=1}^{N_I} f_i}{n} \times 100\%$	

Con las formulas anteriores se obtiene la tabla siguiente:

Tabla 1.12 datos agrupados; intervalos de clase real.

Categoría o Intervalo de Clase Real	**frecuencia absoluta** f_i	**frecuencia relativa** f_{Ri}	**frecuencia porcentual;** $f_{Ri}\%$,	**frecuencia acumulada absoluta** F_i,	**frecuencia acumulada relativa;** F_{Ri}	**frecuencia acumulada porcentual;** $F_{Ri}\%$	**marcas de clase;** m_i
33.5 - 34.5	**1**	**1/40**	**2.5%**	**1**	**1/40**	**2.5%**	**34**
34.5 - 35.5	**2**	**2/40**	**5%**	**3**	**3/40**	**7.5%**	**35**
35.5 - 36.5	**3**	**3/40**	**7.5%**	**6**	**6/40**	**15%**	**36**
36.5 - 37.5	**24**	**24/40**	**60%**	**30**	**30/40**	**75%**	**37**
37.5 - 38.5	**10**	**10/40**	**25%**	**40**	**40/40**	**100%**	**38**
	$n = 40$		**100%**				

A continuación calcularemos las medidas de tendencia central; media aritmética; M_A, mediana; M_E, moda; M_O y las medidas de dispersión: varianza; $Var(x)$ y desviación estándar: DE ,el coeficiente de variación; C_V para datos agrupados en la tabla.

1.4.1 CÁLCULO DE LAS MEDIDAS CENTRALES; LA MEDIA ARITMÉTICA, MEDIANA Y MODAPARA DATOS AGRUPADOS.

Tabla 1.13 fórmula para calcular la **media aritmética.**

media aritmética , M_A $M_A = \dfrac{\sum_{I=1}^{N_I} m_i \cdot f_i}{n}$	Donde n representa el total de datos, m_i la marca de clase y f_i representa la frecuencia absoluta del intervalo de clase i- esimo.

De la tabla podemos observar que la amplitud del intervalo de clase es $A_I = 1$, para el cálculo de la media aritmética, $M_A = \dfrac{\sum m_i f_i}{n}$ solo se necesitan las frecuencias absolutas, las marcas de clase , y la columna de los intervalos de clase real.

Tabla 1.14 columnas de la tabla 1.12 para calcular la media aritmética.

Categoría o Intervalo De Clase Real	frecuencia absoluta f_i	marcas de clase; m_i	$m_i \cdot f_i$
33.5 - 34.5	1	34	34
34.5 - 35.5	2	35	70
35.5 - 36.5	3	36	108
36.5 - 37.5	24	37	888
37.5 - 38.5	10	38	380
	$n = 40$		$\sum_{I=1}^{5} m_i \cdot f_i = 1480$

Es decir: $M_A = \frac{\sum_{I=1}^{N_I} m_i \cdot f_i}{n} = \frac{1\cdot 34 + 2\cdot 35 + 3\cdot 36 + 24\cdot 37 + 10\cdot 38}{40} = \frac{1480}{40} = 37$

Para el cálculo de la mediana, se utilizará la siguiente formula:

Tabla 1.15 fórmula para calcular la Mediana.

Mediana, M_E $M_E = L_{iM_E} + \left[\frac{\frac{n}{2} - F_{AA}}{f_{M_o}} \right] A_I$	L_{iM_E} , representa el límite inferior de clase mediana, n Es el número de datos , F_{AA} la frecuencia acumulada anterior al intervalo de clase de la mediana, f_{M_o} es la frecuencia del intervalo mediano, A_I , es la amplitud del intervalo de clase.

Observamos que se requiere determinar el intervalo mediano, este intervalo será reconocible porque contiene el 50 por ciento de los datos, los datos son en este caso 40, entonces el 50% es 20 datos.

Para obtener la mediana; solo se necesitan las frecuencias absolutas y las frecuencias absolutas acumuladas y la columna de las clases reales.

En la columna de frecuencia acumulada absoluta F_i , en el cuarto intervalo: “36.5 - 37.5”, se tiene el dato ordinal séptimo, octavo,…, treintavo. Por consiguiente el dato veinteavo y veintiunavo le correspondería en el caso de datos no agrupados la categoría de mediana, en esto se basa el criterio para determinar cuál es el intervalo mediano.

Tabla 1.16 columnas de la tabla 1.12 para calcular la mediana.

Categoría o Intervalo de Clase Real	frecuencia absoluta f_i	frecuencia acumulada absoluta F_i ,
33.5 - 34.5	1	1
34.5 - 35.5	2	3
35.5 - 36.5	3	6
36.5 - 37.5 ***Intervalo mediano***	24	30
37.5 - 38.5	10	40
	$n = 40$	

El límite inferior de este intervalo tiene un valor; $L_{iM_E} = 36.5$, la frecuencia del intervalo mediano , podemos encontrarla en la columna de frecuencias absolutas, y es $f_{M_o} = 24$, la amplitud del intervalo es la diferencia de los límites de la clase inferior y superior; $A_I = L_s - L_i$ del mismo intervalo, en este caso se puede elegir cualquiera de los cinco que aparecen en la tabla, por ejempló: $A_I = 37.5 - 36.5 = 1$.

La frecuencia acumulada anterior al intervalo de clase de la mediana F_{AA}, se obtiene de la columna de frecuencia acumulada absoluta F_i, en este caso $F_{AA} = 6$, con esta información sustituimos en:

$$M_E = L_{iM_E} + \left[\frac{\frac{n}{2} - F_{AA}}{f_{M_o}} \right] A_I = 36.5 + \left[\frac{\frac{40}{2} - 6}{24} \right] \cdot 1 = 37.0833$$

Para el cálculo de la moda, utilizaremos la siguiente formula:

Tabla 1.17 fórmula para calcular la Moda.

Moda $M_O = L_{iM_O} + \left[\dfrac{\Delta_1}{\Delta_1 + \Delta_2} \right] A_I$	L_{iM_O} , representa al límite inferior de la clase modal, Δ_1 representa a la diferencia en valor absoluto de la frecuencia de la clase modal y la frecuencia del intervalo premo Δ_2, representa la diferencia en valor absoluto entre la frecuenci modal y la frecuencia del intervalo postmodal. A_I , es la amplitu del intervalo de clase.

Solo se necesitaran la columna de las categorías de clase real y la columna de la frecuencia absoluta.

Para determinar el intervalo modal, observamos la columna de las frecuencias absolutas y determinamos cual es el que tiene mayor frecuencia, es este caso es el cuarto intervalo, y $L_{iM_O} = 36.5$

Tabla 1.18 columnas de la tabla 1.12 para calcular la moda.

Categoría o Intervalo de Clase Real	**Frecuencia absoluta** f_i
33.5 - 34.5	**1**
34.5 - 35.5	**2**
35.5 - 36.5	**3** *Frecuencia premodal*
36.5 - 37.5 *Intervalo modal*	**24** ***Frecuencia modal***
37.5 - 38.5	**10** *Frecuencia postmodal*
	$n = 40$

Para determinar Δ_1, *se toma la diferencia en valor* la diferencia en valor absoluto de la frecuencia de la clase modal y la frecuencia del intervalo premodal,

$\Delta_1 = |24 - 3| = 21$,

para obtener Δ_2, se toma la diferencia en valor absoluto entre la frecuencia modal y la frecuencia del intervalo postmodal:

$\Delta_2 = |24 - 10| = 14$,

la amplitud del intervalo es la diferencia de los límites de la clase inferior y superior; $A_I = L_s - L_i$ del mismo intervalo, en este caso se puede elegir cualquiera de los cinco que aparecen en la tabla, por ejempló: $A_I = 37.5 - 36.5 = 1$. Con estos datos, se sustituye en la fórmula:

$$M_O = L_{iM_O} + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2}\right] A_I = 36.5 + \left[\frac{21}{21 + 14}\right] \cdot 1 = 37.1$$

Es conveniente comparar estos resultados obtenidos usando la tabla de datos agrupados y determinar el margen de error por agrupamiento de datos.

En la siguiente tabla contiene los valores que se han obtenido para los datos sin agrupar y para los datos agrupados.

Tabla 1.19 comparación de resultados de las medidas centrales; media aritmética, mediana y moda.**Para datos no agrupados y datos agrupados.**

	Datos Sin agrupar	Datos agrupados
Media aritmética	**37**	**37**
Mediana	**37**	**37.0833**

moda	**37**	**37.1**

ACTIVIDAD 25

Determine el coeficiente de variación porcentual de las medidas centrales obtenidas por medio de procedimientos de datos sin agrupar y por medio de procedimientos de datos agrupados. Comente los resultados obtenidos mediante el coeficiente de variación porcentual.

Solución.

	Datos Sin agrupar	Datos agrupados	Media aritmética	Desviación estándar	Coeficiente De Variación Porcentual $CV\% = \frac{DE}{M_A} \times 100\%$
Media aritmética	**37**	**37**	**37**	**0**	**0 %**
Mediana	**37**	**37.0833**	**37.0416**	**0.0208**	**0.0562 %**
moda	**37**	**37.1**	**37.05**	**0.05**	**0.125 %**

Comentario:

el coeficiente de variación es muy pequeño, en el mayor de los casos no sobrepasa el 0,2 %, lo cual indica que el agrupamiento de datos fue bastante bueno.

1.4. 2 CÁLCULO DE LAS MEDIDAS DE DISPERSIÓN: VARIANZA, DESVIACIÓN ESTÁNDAR Y COEFICIENTE DE VARIACIÓN PARA DATOS AGRUPADOS.

Para el cálculo de la varianza, desviación estándar y coeficiente de variación, se hará uso de las siguientes formulas:

Tabla 1.20 algunas fórmulas para medidas de dispersión, para datos agrupados.

Varianza $Var(x)=\frac{\sum_{I=1}^{N_I}\left[m_i-M_A\right]^2 f_i}{n}$	Donde n representa el total de datos, m_i la marca de clase y f_i representa la frecuencia absoluta del intervalo de clase i- esimo, M_A Representa la media Aritmética.
Desviación estándar $DE=\sqrt{\frac{\sum_{I=1}^{N_I}(m_i-M_A)^2}{n}}$	Es decir si se dispone de la varianza, la raíz cuadr de esta es la deviación estándar, igual que con lo datos no agrupados.
Coeficiente de variación $CV=\frac{DE}{M_A}$	Es decir el coeficiente de variación es el cociente la deviación estándar entre La media aritmética, igual que en el caso de datos no agrupados.

Para el cálculo de la varianza; $Var(x)$ necesitamos los datos de las columnas; intervalo de clases, la frecuencia absoluta y marcas de clase, y agregaremos otra columna; $(m_i - M_A)^2$, $(m_i - M_A)^2 \cdot f_i$,además se requiere el valor de la media aritmética: $M_A = 37$

Tabla 1.21 columnas de la tabla 1.12 para calcular la varianza.

Categoría o Interv. de Clase Real	**frecuencia absoluta** f_i	**marcas de clase** m_i	$(m_i - M_A)^2$	$(m_i - M_A)^2 \cdot f_i$
33.5 - 34.5	**1**	**34**	**(34-37)^2 = 9**	**9(1) = 9**
34.5 - 35.5	**2**	**35**	**(35-37)^2 = 4**	**4(2) = 8**
35.5 - 36.5	**3**	**36**	**(36-37)^2 = 1**	**1(3) = 3**
36.5 - 37.5	**24**	**37**	**(37-37)^2 = 0**	**0(24) = 0**
37.5 - 38.5	**10**	**38**	**(38-37)^2 = 1**	**1(10) = 10**
	$n = 40$			$\sum_{I=1}^{5}(m_i - M_A)^2 \cdot f_i = 30$

Sustituyendo valores en la expresión de la varianza:

$$Var(x) = \frac{\sum_{I=1}^{N_I}\left[m_i - M_A\right]^2 f_i}{n} = \frac{(34-37)^2\cdot 1 + (35-37)^2\cdot 2 + (36-37)^2\cdot 3 + (37-37)^2\cdot 24 + (38-37)^2\cdot 10}{40} = \frac{30}{40} = 0.75$$

Para el cálculo de la desviación estándar:

$$DE = \sqrt{\frac{\sum_{I=1}^{N_I}(m_i - M_A)^2}{n}} = \sqrt{0.75} = 0.8660$$

Para el cálculo del coeficiente de variación:

$$CV = \frac{DE}{M_A} = \frac{0.8660}{37} = 0.0234$$

¿Cuál será el valor de la varianza y Varianza muestral?, ¿Si se conoce una de ellas, será posible encontrar la otra?

De los datos anteriores, se sabe que la varianza poblacional tiene un valor de 0.75, es decir:

$$Var(x) = \frac{\sum_{I=1}^{N_I} \left[m_i - M_A \right]^2 f_i}{n} = 0.75$$

la varianza poblacional y la varianza muestral solo difieren en que el divisor de la varianza poblacional es el número de datos "n", y el divisor de la varianza muestral es "$n-1$"

$$Var(x)_m = \frac{\sum_{I=1}^{N_I} \left[m_i - M_A \right]^2 f_i}{n-1}$$

Con los datos disponibles, para pasar de una a otra solo se necesita conocer

$$\sum_{I=1}^{N_I} \left[m_i - M_A \right]^2 f_i$$

Es decir:

$$\sum_{I=1}^{N_I} \left[m_i - M_A \right]^2 f_i =$$

$$= n \cdot (0.75) = 40 \cdot (0.75) = 30$$

Por lo que la varianza muestral:

$$Var(x)_m = \frac{\sum_{I=1}^{N_I} \left[m_i - M_A \right]^2 f_i}{n-1} =$$

$$= \frac{30}{40-1} = 0.7692$$

¿Cuál será el valor de la desviación estándar y desviación estándar muestral?, Si ya se conocen las varianzas, tanto poblacional como muestral entonces, es posible deducir la otra.

Tabla 1.22 comparación de fórmulas para la población y para la muestra.

varianza poblacional

$$Var(x) = \frac{\sum_{I=1}^{N_I} (m_i - M_A)^2 \cdot f_i}{n} = 0.75$$

$$DE = \sqrt{\frac{\sum_{I=1}^{N_I} (m_i - M_A)^2 \cdot f_i}{n}} = \sqrt{0.75} =$$
$$= 0.8660$$

varianza muestral

$$Var(x)_m = \frac{\sum_{I=1}^{N_I} (m_i - M_A)^2 \cdot f_i}{n-1} = 0.7692$$

$$DE_m = \sqrt{\frac{\sum_{I=1}^{N_I} (m_i - M_A)^2 \cdot f_i}{n-1}} = \sqrt{0.7682} =$$
$$= 0.8765$$

A continuación se deduce la formula reducida de la varianza para datos agrupados.

La definición operativa de varianza poblacional de datos agrupados es:	$Var(x) = \frac{\sum_{I=1}^{N_I} (m_i - M_A)^2 \cdot f_i}{n}$
Donde n representa el total de datos, m_i la marca de clase y	f_i ,representa la frecuencia absoluta del intervalo de clase i- esimo, M_A Representa la mediaaritmética.
Desarrollando el binomio	$Var(x) = \frac{\sum_{I=1}^{N_I} \left[m_i^2 - 2M_A m_i + M^2{}_A \right] \cdot f_i}{n}$

cuadrado	
Reescribiendo:	$Var(x)=\frac{1}{n}\sum_{I=1}^{N_I}\left[m_i^2-2M_A m_i+M^2{}_A\right]\cdot f_i$
Distribuyendo f_i	$Var(x)=\frac{1}{n}\sum_{I=1}^{N_I}\left[m_i^2\cdot f_i-2M_A m_i\cdot f_i+M^2{}_A\cdot f_i\right]$
Aplicando la suma a cada termino:	$Var(x)=\frac{1}{n}\left[\sum_{I=1}^{N_I}m_i^2\cdot f_i+\sum_{I=1}^{N_I}[-2M_A m_i\cdot f_i]+\sum_{I=1}^{N_I}M^2{}_A\cdot f_i\right]$
Aplicando ley distributiva para $\frac{1}{n}$,	$Var(x)=\frac{1}{n}\sum_{I=1}^{N_I}m_i^2\cdot f_i+\frac{1}{n}\sum_{I=1}^{N_I}[-2M_A m_i\cdot f_i]+\frac{1}{n}\sum_{I=1}^{N_I}M^2{}_A\cdot f_i$
Aplicando propiedades de la Suma.	$Var(x)=\frac{1}{n}\cdot\sum_{I=1}^{N_I}m_i^2\cdot f_i-2M_A\frac{1}{n}\cdot\sum_{I=1}^{N_I}m_i\cdot f+\frac{1}{n}\cdot n\cdot M^2{}_A$
Las constantes pueden salir fuera del símbolo suma.	La suma de las frecuencias es el total de datos: $\sum_{I=1}^{N_I}f_i=n$
el termino $\frac{1}{n}\cdot\sum_{i=1}^{n}m_i\cdot f_i$	Es precisamente la media aritmética de datos Agrupados: M_A

Simplificando	$Var(x)=\frac{1}{n}\cdot\sum_{I=1}^{N_I}m_i^{2}\cdot f_i-2M^2{}_A+M^2{}_A$
Reduciendo, tenemos :	$Var(x)=\frac{1}{n}\cdot\sum_{I=1}^{N_I}m_i^{2}\cdot f_i-M^2{}_A$

Tabla 1.23 fórmula para calcular la varianza

formula reducida de la varianza, para datos agrupados
$Var(x)=\frac{1}{n}\cdot\sum_{I=1}^{N_I}m_i^{2}\cdot f_i-M^2{}_A$

Para revisar la utilidad y mostrar las ventajas que presenta, se utilizaran los datos de la tabla de datos agrupados del problema 5, los cuales previamente se han obtenido , se necesitan las columnas, intervalo de clases, la frecuencia absoluta y marcas de clase, y agregaremos otra columna; , , además se requiere el valor de la media aritmética:

Tabla 1.24 columnas de la tabla 1.12 para calcular la varianza reducida.

Categoría o Intervalo de Clase Real	frecuencia absoluta f_i	Marcas de Clase m_i	m_i^2	$m_i^2\cdot f_i$
33.5 - 34.5	1	34	$34^2 = 1156$	1156(1) = 1156
34.5 - 35.5	2	35	$35^2 = 1225$	1225(2) = 2450
35.5 - 36.5	3	36	$36^2 = 1296$	1296(3) = 3888

36.5 - 37.5	24	37	$37^2 = 1369$	1369(24) = 32856
37.5 - 38.5	10	38	$38^2 = 1444$	1444(10) = 14440
	$n = 40$			$\sum_{I=1}^{5} m_i^2 \cdot f_i = 54790$

Aplicando la fórmula de la varianza reducida con los datos de la tabla, obtenemos varianza:

$Var(x) = \frac{1}{n} \cdot \sum_{i=1}^{n} m_i^2 \cdot f_i - M^2{}_A = \frac{1}{40} \cdot (54790) - (37)^2 = 0.75$, el cual es el mismo valor encontrado

con la definición operativa de la varianza formula $Var(x) = \frac{\sum_{I=1}^{N_I} (m_i - M_A)^2 \cdot f_i}{n} = 0.75$

ACTIVIDAD 26

Determine los valores de la varianza muestral, desviación estándar poblacional y la desviación estándar muestral. Si se conoce que la varianza poblacional tiene un valor de 3.6, y que esta fue estimada a partir de 100 datos.

Solución.

A partir de la definición varian poblacional:	$Var(x) = \frac{\sum_{I=1}^{N_I} (m_i - M_A)^2 \cdot f_i}{n}$
Se tiene que:	$\frac{\sum_{I=1}^{N_I} (m_i - M_A)^2 \cdot f_i}{100} = 3.6$

Despejando: $\sum_{i=1}^{n}(m_i - M_A)^2 \cdot f_i$	$\sum_{I=1}^{N_I}(m_i - M_A)^2 \cdot f_i = 3.6 \cdot (100) = 360$
La varianza muestral es :	$Var(x)_m = \dfrac{\sum_{I=1}^{N_I}\left[m_i - M_A\right]^2 f_i}{n-1} = \dfrac{360}{100-1} = 3.6363$
La desviación estándar muestra es:	$DE_m = \sqrt{\dfrac{\sum_{i=1}^{N_I}(m_i - M_A)^2 \cdot f_i}{n-1}} = \sqrt{3.6363} =$ $= 1.9069$
La desviación estándar poblaci es:	$DE = \sqrt{\dfrac{\sum_{i=1}^{N_I}(m_i - M_A)^2 \cdot f_i}{n}} = \sqrt{3.6} =$ $= 1.8974$

ACTIVIDAD 27

Determine los valores de la varianza poblacional, desviación estándar poblacional y la varianza muestral. Si se conoce que la desviación estándar muestral, y ésta tiene un valor de 6.28, y el total de datos en una muestra donde fue calculado este valor fue de 30 datos.

Solución.

A partir de la definición de la desviación estándar muestral	$DE_m = \sqrt{\dfrac{\sum_{i=1}^{N_I} (m_i - M_A)^2 \cdot f_i}{n-1}}$
Se tiene que:	$DE_m = \sqrt{\dfrac{\sum_{i=1}^{N_I} (m_i - M_A)^2 \cdot f_i}{30-1}} = 6.28$
Despejando: $\sum_{i=1}^{n} (m_i - M_A)^2 \cdot f_i$	$Var(x)_m = \dfrac{\sum_{i=1}^{N_I} (m_i - M_A)^2 \cdot f_i}{30-1} = (6.28)^2 = 39.4384$
Es decir, la varianza muestral es:	$Var(x)_m = 39.4384$
Despejando: $\sum_{i=1}^{n} (m_i - M_A)^2 \cdot f_i$	$\sum_{I=1}^{N_I} (m_i - M_A)^2 \cdot f_i = (39.4384)\cdot(29) = 1143.7136$
La varianza poblacional es :	$Var(x) = \dfrac{\sum_{I=1}^{N_I} (m_i - M_A)^2 \cdot f_i}{n}$
Sustituyendo datos:	$Var(x) = \dfrac{\sum_{I=1}^{N_I} (m_i - M_A)^2 \cdot f_i}{n} = \dfrac{1143.7136}{30} = 38.1238$
Es decir la varianza poblacional tiene el valor:	$Var(x) = 38.1238$

La desviación estándar poblaci :	$DE=\sqrt{\dfrac{\sum_{I=1}^{N_I}(m_i-M_A)^2\cdot f_i}{n}}=\sqrt{38.1238}=$ $=6.1744$

A continuación se deduce una relación entre la varianza poblacional y muestral:

La varianza poblacional es:	$Var(x)=\dfrac{\sum_{I=1}^{N_I}(m_i-M_A)^2\cdot f_i}{n}$
Despejando: $\sum_{i=1}^{N_I}(m_i-M_A)^2\cdot f_i$	$\sum_{I=1}^{N_I}(m_i-M_A)^2\cdot f_i=n\cdot Var(x)\ldots\quad(1)$
Por otra parte la varianza muestral es:	$Var(x)_m=\dfrac{\sum_{I=1}^{N_I}(m_i-M_A)^2\cdot f_i}{n-1}$
Despejando: $\sum_{i=1}^{N_I}(m_i-M_A)^2\cdot f_i$	$\sum_{I=1}^{N_I}(m_i-M_A)^2\cdot f_i=(n-1)Var(x)_m\ldots\quad(2)$
Igualando las expresiones (1) y (2):	$\sum_{I=1}^{N_I}(m_i-M_A)^2\cdot f_i=n\cdot Var(x)=(n-1)Var(x)_m$
Es decir:	$n\cdot Var(x)=(n-1)Var(x)_m$
Entonces: $Var(x)=\dfrac{(n-1)}{n}Var(x)_m$	o bien: $Var(x)_m=\dfrac{n}{(n-1)}\cdot Var(x)$

ACTIVIDAD 28

a. Verifique que si varianza poblacional tiene un valor de 3.6, y que esta fue estimada a partir de 100 datos, la desviación estándar muestral es 1.9069

b. Verifique que si desviación estándar muestral tiene un valor de 6.28, y que esta fue estimada a partir de 30 datos, la varianza a poblacional es 38.1238

Solución.

$a.\ si:\quad Var(x)_m = 3.6$	$Var(x)_m = \frac{100}{(100-1)}\cdot(3.6) = 3.6363$
La desviación estándar muestral:	$DE_m = \sqrt[2]{Var(x)_m} = \sqrt{3.6363} = 1.9069$
$b.\ si:\quad DE_m = 6.28$	$Var(x) = \frac{(n-1)}{n}Var(x)_m =$ $= \frac{(n-1)}{n}(DE_m)^2 = \frac{(30-1)}{30}(6.28)^2 =$ $= 38.1238$

ACTIVIDAD 29

Considere la siguiente tabla de datos agrupados. Determine lo que se indica en los incisos.

Tabla 1.25 datos; intervalos y frecuencias absolutas.

Categoría o Intervalo de Clase Real	**frecuencia absoluta** f_i
-1.01 - 0.49	**30**
0.49 - 1,99	**24**
1.99 - 3.49	**6**
3.49 - 4.99	**10**
4.99 - 6.49	**12**
6.49 - 7.99	**8**
	$n = 90$

a. las marcas de clase m_i

b. la frecuencia absoluta acumulada F_i

c. la longitud del intervalo de clase real, I_{CR}

Solución.

Categoría o Intervalo de Clase Real	**frecuencia absoluta** f_i	**frecuencia acumulada absoluta** F_i **,**	**marcas de clase;** m_i
-1.01 - 0.49	**30**	**30**	**-0.26**
0.49 - 1,99	**24**	**54**	**1.24**

1.99 - 3.49	**6**	**60**	**2.74**
3.49 - 4.99	**10**	**70**	**4.24**
4.99 - 6.49	**12**	**82**	**5.74**
6.49 - 7.99	**8**	**90**	**7.24**
la longitud del intervalo de clase real, $A_I = 1.5$	$n = 90$		

ACTIVIDAD 30

Considere la siguiente tabla, determine la media aritmética, la mediana y la moda de datos agrupados.

Tabla 1.26 datos para calcular medidas centrales.

Categoría o Intervalo de Clase Real	**frecuencia absoluta** f_i	**frecuencia acumulada absoluta** F_i **,**	**marcas de clase;** m_i
-1.01 - 0.49	**30**	**30**	**-0.26**
0.49 - 1,99	**24**	**54**	**1.24**
1.99 - 3.49	**6**	**60**	**2.74**

3.49 - 4.99	**10**	**70**	**4.24**
4.99 - 6.49	**12**	**82**	**5.74**
6.49 - 7.99	**8**	**90**	**7.24**

Solución.

Media Aritmética

$$M_A = \frac{\sum_{I=1}^{N_I} m_i \cdot f_i}{n} =$$

$$= \frac{(-0.26)\cdot(30)+(1.24)\cdot(24)+(2.74)\cdot(6)+(4.24)\cdot(10)+(5.74)\cdot(12)+(7.24)\cdot(8)}{90} =$$

$$= \frac{2076}{90} = 2.3067$$

Mediana, el intervalo mediano es : **0.49 - 1,99,** $n = 90$, $f_{M_E} = 24$

, $A_I = 1.5$, $F_{AA} = 30$

$$M_E = L_{iM_E} + \left[\frac{\frac{n}{2} - F_{AA}}{f_{M_E}} \right] A_I = (0.49) + \left[\frac{\frac{(90)}{2} - (30)}{(24)} \right] (1.5) = 1.4275$$

Moda, el intervalo modal es: **-1.01 - 0.49,** $\Delta_1 = 6$, $\Delta_2 = 18$ $A_I = 1.5$

$$M_O = L_{iM_O} + \left[\frac{\Delta_1}{\Delta_1 + \Delta_2} \right] A_I = (-1.01) + \left[\frac{(6)}{(6)+(18)} \right] (1.5) = -0.635$$

ACTIVIDAD 31

Considere la siguiente tabla, determine la varianza, la desviación estándar poblacionales y muestrales, el coeficiente de variación y el coeficiente de variación porcentual de datos agrupados.

Tabla 1.27 datos para calcular medidas de dispersión.

Categoría o Intervalo de Clase Real	**Frecuencia absoluta** f_i	**marcas de clase;** m_i
-1.01 - 0.49	**30**	**-0.26**
0.49 - 1,99	**24**	**1.24**
1.99 - 3.49	**6**	**2.74**
3.49 - 4.99	**10**	**4.24**
4.99 - 6.49	**12**	**5.74**

Universidad Nacional Autónoma de México UNAM — Portal académico Colegio de Ciencias y Humanidades

6.49 - 7.99	**8**	**7.24**

Solución.

Calculando la formula reducida de la varianza.

Clase Real	f_i	m_i	m_i	m_i f_i
-1.01 - 0.49	**30**	**-0.26**	**$(-0.26)^2 = 0.0676$**	**0.0676(30) = 2.028**
0.49 - 1,99	**24**	**1.24**	**$(1.24)^2 = 1.5376$**	**1.5376(24) = 36.9024**
1.99 - 3.49	**6**	**2.74**	**$(2.74)^2 = 7.5076$**	**7.5076(6) = 45.0456**
3.49 - 4.99	**10**	**4.24**	**$(4.24)^2 = 17.9776$**	**17.9776(10) = 179.776**
4.99 - 6.49	**12**	**5.74**	**$(5.74)^2 = 32.9476$**	**32.9476(12) = 395.3712**
6.49 - 7.99	**8** $n = 90$	**7.24**	**$(7.24)^2 = 52.4176$**	**52.4176(8) = 419.3408** $\sum_{I=1}^{5} m_i^2 \cdot f_i = 1078.464$

Usando la formula reducida de la varianza: $Var(x) = \frac{1}{n} \cdot \sum_{I=1}^{N_I} m_i^2 \cdot f_i - M_A^2 =$ $= \frac{1}{90} \cdot (1078.464)_i - (2.3067)^2 =$ $= 6.6621$
La desviación estándar poblacional: $DE = \sqrt{Var(x)} = \sqrt{6.6621} = 2.5811$
Para determinar el valor de la varianza muestral: $Var(x)_m = \frac{n}{(n-1)} \cdot Var(x) = \frac{90}{(90-1)} \cdot (6.6621) = 6.7369$
La desviación estándar muestral : $DE_m = \sqrt{Var_m(x)} = \sqrt{6.7369} = 2.5955$

1.5 MEDIDAS DE POSICIÓN: LOS CUANTILES; CUARTILES, DECILES Y PORCENTILES DE DATOS AGRUPADOS.

Se ha comentado anteriormente en la sección 1.3.1 que los llamados cuantiles, no son otra cosa que marcas (medidas e posición)que establecen divisiones, usualmente en partes iguales, por ejemplo se les llama cuartiles si se divide en cuatro partes el conjunto de datos, deciles si el total de datos se divide entre diez o porcentiles si el total de los datos se divide entre 100.

La utilidad de estas medidas estriba entre otras cosas, en que permite determinar si existe simetría de la distribución de datos unimodal (con un solo máximo), tales como el sesgo y la curtosis (grado de agudeza) en las distribuciones que tienen un solo máximo

El término "cuantil" fue usado por primera vez por Maurice George Kendall en 1940. El cuantil marca un corte de modo que una proporción p de valores de la población es menor o igual que la proporción de dicho cuantil. Recordando lo ya visto en la sección 1.3.1, tenemos:

a. Los Cuartiles; Q_1, Q_2, Q_3 que dividen a la distribución en cuatro partes, corresponden a la división de la población de datos en el 25%, 50% y 75%.

b. Los **Deciles**, $D_1, D_2, D_3..., D_9$ que dividen a la distribución en diez partes; es decir en el 10%, 20%,…90%.

c. Los **Percentiles**, $P_1,..., P_{99}$ que dividen a la distribución en cien partes, es decir; 1%, 2%,…,99%

Tabla 1.28 fórmulas para calcular medidas de centrales.

Consideremos la fórmula de la mediana para datos agrupados, la cual coincide con el cuartil número dos,

$$M_E = L_{iM_E} + \left[\frac{\frac{n}{2} - F_{AA}}{f_{M_o}} \right] A_I$$

ésta la podemos adaptar para realizar los cálculos de cualquiera de cuantiles C_k; cuartiles deciles o porcentiles, simplemente la expresamos para realizar cálculos de la información dividida en 100 partes:

$$C_k = L_k + \left[\frac{n \cdot k - F_{AAk}}{f_k} \right] A_I$$

Fórmula para el cálculo de cuantiles:

$$C_k = L_k + \left[\frac{n \cdot k - F_{AAk}}{f_k} \right] A_I$$

C_k : cuantil k-esimo
L_k : limite inferior de la categoria correspondiente a la proporción k
n : numero de datos
k : proporción de datos
F_{AAK} : frecuencia acumulada absoluta anterior a la categoria k
f_k : frecuencia de la categoria k

A continuación se muestra como se utiliza la formula, pero se requiere una tabla de datos agrupados, considere la tabla 1.12, que se indica abajo:

Categoría o Intervalo de Clase Real	**frecuencia absoluta f_i**	**frecuencia acumulada absolu** **F_i,**
33.5 - 34.5	**1**	**1**
34.5 - 35.5	**2**	**3**
35.5 - 36.5	**3**	**6**
36.5 - 37.5	**24**	**30**
37.5 - 38.5	**10**	**40**
	$n = 40$	

Supóngase que se desea calcular el cuartil 3, Q_3 esta marca le corresponde al 75 % ($k = 0.75$) de los datos, de acuerdo a la tabla 1.12, el total de datos es $n = 40$.

la amplitud del intervalo se puede obtener restando los limites superior e inferior de cualquier categoría o intervalo: **33.5 - 34.5,** y este es igual a **1.**

La frecuencia del intervalo que contiene el 75% de los datos, se puede obtener a través de: $n \cdot k = (40) \cdot (0.75) = 30$, es decir tenemos que buscar en que intervalo cae el ordinal 30, observamos que está en el cuarto intervalo, la frecuencia que le corresponde al intervalo es $f_k = 24$.

Mientras que la frecuencia acumulada absoluta del intervalo anterior es: $F_{AAk} = 6$. Sustituyendo datos:

$$C_k = L_k + \left[\frac{n \cdot k - F_{AAk}}{f_k}\right] A_I \qquad C_{0.75} = (36.5) + \left[\frac{40 \cdot (0.75) - (6)}{(24)}\right](1) = 37.5$$

ACTIVIDAD 32

Determine los valores de los deciles D_1, D_3, D_9 considere los datos de la tabla 1.12,

Solución.

a.

$n \cdot k = (40) \cdot (0.10) = 4$, buscamos el ordinal 4, cae en el segundo Intervalo:

34.5 - 35.5, $f_k = 2$, $F_{AAk} = 1$, $A_I = 1$, $n = 40$,

sustituyendo datos:

$$D_1 = C_{0.10} = (34.5) + \left[\frac{(40) \cdot (0.10) - (1)}{(2)}\right](1) = 36$$

b.

$n \cdot k = (40) \cdot (0.30) = 12$, buscamos el ordinal 12, cae en el cuarto Intervalo:

36.5 - 37.5, $f_k = 24$, $F_{AAk} = 6$, $A_I = 1$, $n = 40$,

sustituyendo datos:

$$D_3 = C_{0.30} = (36.5) + \left[\frac{(40) \cdot (0.30) - (6)}{(24)}\right](1) = 36.75$$

c.

$n \cdot k = (40) \cdot (0.90) = 36$, buscamos el ordinal 12, cae en el quinto Intervalo:

37.5 - 38.5, $f_k = 10$, $F_{AAk} = 30$, $A_I = 1$, $n = 40$,

sustituyendo datos:

$$D_9 = C_{0.90} = (37.5) + \left[\frac{(40) \cdot (0.90) - (30)}{(10)} \right] (1) = 38.10$$

ACTIVIDAD 33

Determine los valores de los porcentiles P_{36}, P_{66}, P_{87} considere los datos de la tabla 1.12,

Solución.

a.

$n \cdot k = (40) \cdot (0.36) = 14.4 \approx 15$, buscamos el ordinal 15, cae en el cuarto Intervalo:

36.5 - 37.5, $f_k = 24$, $F_{AAk} = 6$, $A_I = 1$, $n = 40$,

sustituyendo datos:

$$P_{36} = C_{0.36} = (36.5) + \left[\frac{(40) \cdot (0.36) - (6)}{(24)} \right] (1) = 36.85$$

b.

$n \cdot k = (40) \cdot (0.66) = 26.4 \approx 25$, buscamos el ordinal 25, cae en el cuarto Intervalo:

36.5 - 37.5, $f_k = 24$, $F_{AAk} = 6$, $A_I = 1$, $n = 40$,

sustituyendo datos:

$$P_{66} = C_{0.66} = (36.5) + \left[\frac{(40) \cdot (0.66) - (6)}{(24)} \right] (1) = 37.35$$

c.

$n \cdot k = (40) \cdot (0.87) = 34.8 \approx 35$, buscamos el ordinal 35, cae en el quinto Intervalo:

37.5 - 38.5, $f_k = 10$, $F_{AAk} = 30$, $A_I = 1$, $n = 40$,

sustituyendo datos:

$$P_{87} = C_{0.87} = (37.5) + \left[\frac{(40) \cdot (0.87) - (30)}{(10)} \right] (1) = 37.98$$

1.6 LAS GRÁFICAS DE DATOS ESTADÍSTICOS (datos agrupados). Gráfica de barras, polígono de frecuencias, curva de frecuencias gráfica circular (o de pastel o pay), ojiva (creciente), diagrama de caja y bigote.

1.6.1 Grafica de barras

Se construye marcando en el eje horizontal los limites inferior y superior de los intervalos de clase, y en el eje vertical las frecuencias absolutas (o las relativas), en cualquiera de los casos se debe indicar mediante símbolos que se está representando. Las barras suelen ser verticales, pero también algunas personan suelen dibujarlos horizontales.

En la dirección electrónica http://www.disfrutalasmatematicas.com/graficos/grafico-barras.html, se encuentra un applet que te ayuda a construir dicha gráfica.

Con la gráfica de barras se pueden construir otros diagramas, tales como el polígono de frecuencias, o la curva de frecuencias.

Tabla 1.29 columnas de la tabla 1.12 para construir la gráfica de barras.

Categoría o Intervalo de Clase Real	Frecuencia absoluta f_i
33.5 - 34.5	**1**
34.5 - 35.5	**2**
35.5 - 36.5	**3**
36.5 - 37.5	**24**
37.5 - 38.5	**10**
	$n = 40$

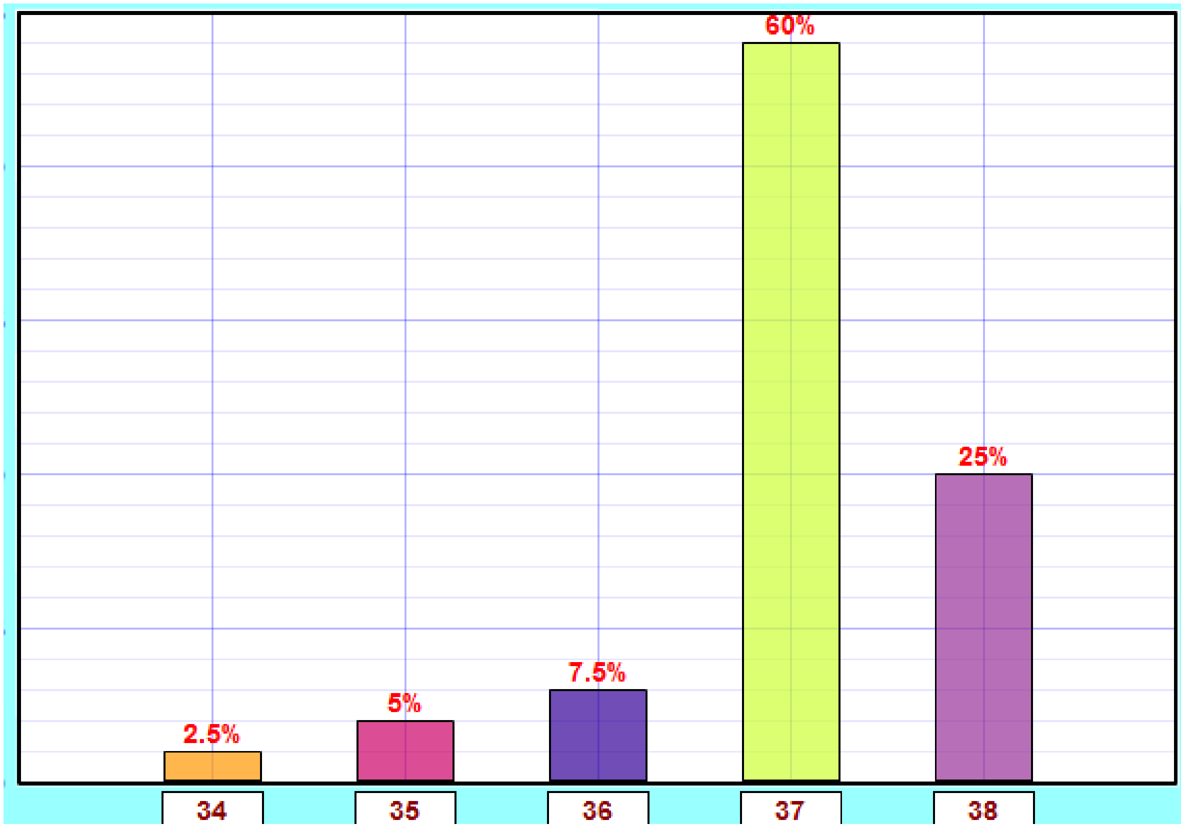


Figura 1.1 diagrama de barras.

1.6.1.1 Los Intervalos de clase.

En determinadas ocasiones es necesario construir un nuevo tipo de intervalo, parecidos a los intervalos de clase real, que ya han sido utilizados anteriormente. También el objetivo es ordenar los datos, pero se busca eliminar ambigüedades. , por ejemplo, pensemos que tenemos disponibles los siguientes datos: 34.0, 34.1, 34, 34.5, 35.9, 36.1, 37.5, 37.6, 38.0, 38.0, el registro disponible es el que se muestra abajo, el dato 34.5 deberá estar en la categoría: **33.5 - 34.5** o en la categoría: **34.5 - 35.5**

Tabla 1.30 intervalos de clase

Categoría o Intervalo de Clase Real	Datos en las categorías	Frecuencia absoluta f_i
33.50 - 34.50	34.0, 34.1, 34.0	**3 ¿ ?**
34.50 - 35.50		**0 ¿ ?**
35.50 - 36.50	35.9, 36.1	**2**
36.50 - 37.50		**0**
37.50 - 38.50	37.5, 37.6, 38.0, 38.0,	**4**
		$n = 10$

Se busca que los nuevos intervalos tengan las siguientes propiedades:

a. Exhaustivos, que contengan a todos los datos obtenidos o disponibles.

b. Excluyentes, Con la finalidad de que un dato caiga en una sola categoría.

La mínima unidad medible, en los datos proporcionados es $U = 0.1$. Dividiendo entre dos y sumando al extremo izquierdo esta cantidad y al extremo derecho de cada intervalo.se tiene:

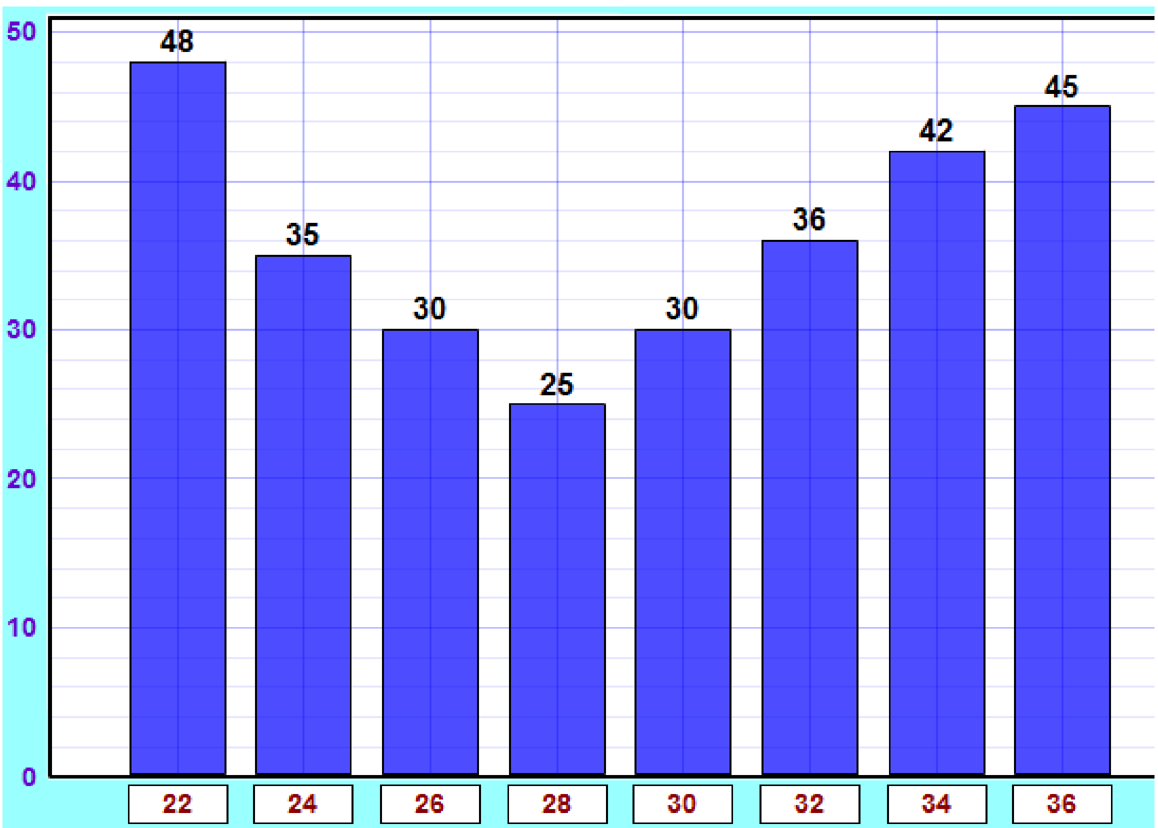
Tabla 1.31 intervalos de clase e intervalos de clase real

Categoría o Intervalo de Clase Real	Categoría o Intervalo de Clase	Datos en las categorías	Frecuencia absol f_i
33.50 - 34.50	**33.55 - 34.45**	34.0, 34.1, 34.0	**3**
34.50 - 35.50	**34.55 - 35.45**	37.5	**1**
35.50 - 36.50	**35.55 - 36.45**	35.9, 36.1	**2**
36.50 - 37.50	**36.55 - 37.45**		**0**
37.50 - 38.50	**37.55 - 38.45**	37.5, 37.6, 38.0, 38.	**4**

			$n=10$

ACTIVIDAD 34

De acuerdo a la información mostrada en la gráfica de barra, completa la distribución de probabilidad correspondiente.



Solución.

Categoría o Intervalo de Clase Real	**Categoría o Intervalo de Clase**	**Frecuencia absoluta f_i**	**Marca de clase m_i**
21-23	**21.5 - 22.9**	**48**	**22**
23-25	**23.0- 24.9**	**35**	**24**
25-27	**25.0 - 26.9**	**30**	**26**
27-29	**27.0 - 37.5**	**25**	**28**
29-31	**29.0 - 38.5**	**30**	**30**
31-33	**31.0 - 32.8**	**36**	**32**
33-35	**33.0 - 34.5**	**42**	**34**
35-37	**35.0 - 36.9**	**45**	**36**
		$n = 291$	

1.6.2 Grafica circular (o de pastel o pay),

En la construcción del diagrama circular, se debe considerar que al círculo le corresponden 360 grados, y que cada a cada categoría le corresponden una cierta cantidad de grados en el círculo.

Usando una regla de tres,se relacionan con los porcentajes que corresponden a cada categoría o intervalo, con el Ángulo, de forma que el al porcentaje de los datos proporcional al ángulo del sector circular.

Se acostumbra a colorear cada sector del círculo con colores distintos e indicar a que intervalo representa.

Para la construcción de la gráfica circular (o de pastel o pay), se deben calcular todos los ángulos correspondientes a todas las categorías. Para esto consideremos la proporción o regla de tres.

Tabla 1.32 intervalos de clase e intervalos de clase real

El ángulo θ, del sector circular guarda relación con el porcentaje de datos del intervalo	$f_{Ri}\%$
El ángulo 360 grados del círculo, guarda relación con el porcentaje de datos del intervalo	100%

Como ejemplo consideremos la siguiente distribución de datos:

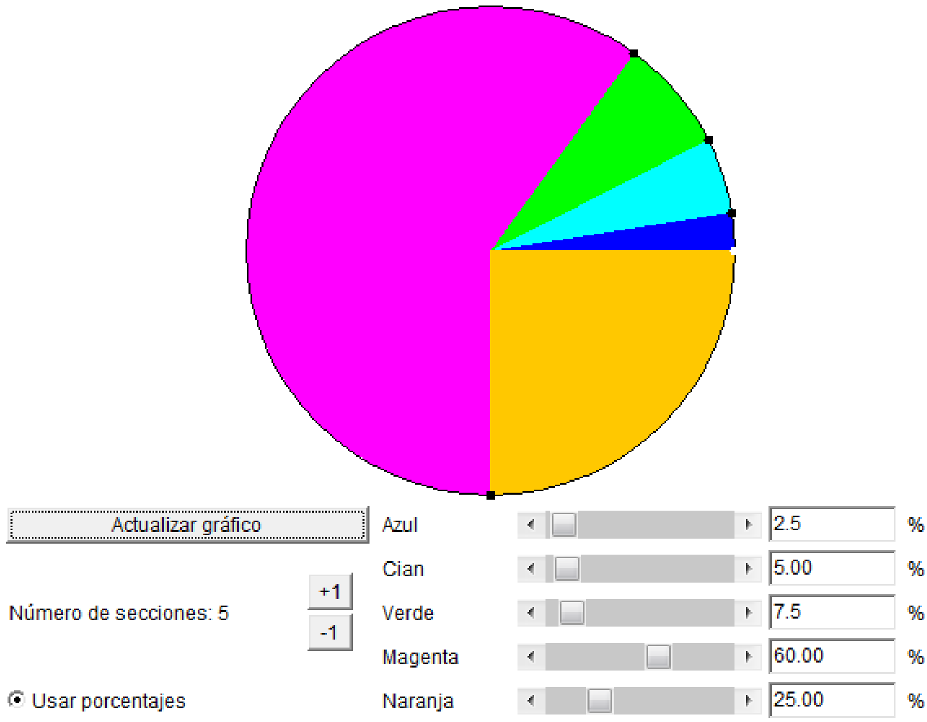
Tabla 1.33 intervalos de clase e intervalos de clase real

Categoría o Intervalo de Clase Real	**Frecuencia absoluta** f_i
33.5 - 34.5	**1**
34.5 - 35.5	**2**
35.5 - 36.5	**3**
36.5 - 37.5	**24**
37.5 - 38.5	**10**
	$n = 40$

Entonces: $\dfrac{\theta}{f_{Ri}\%} = \dfrac{360}{100\%}$, despejando el ángulo del sector: $\theta = \dfrac{360}{100} f_{Ri}$, usando esta relación tenemos:

Tabla 1.34 columnas de la tabla 1.12 para construir la gráfica de barras.

Intervalo de Clase Real	**Color**	**frecuencia porcentual; $f_{Ri}\%$,**	**Angulo Correspondiente** $\theta = \dfrac{360}{100} f_{Ri}$
33.5 - 34.5	**Azul**	**2.5%**	$\theta_1 = \dfrac{360}{100}(2.5) = 3.6(2.5) = 9^\circ$
34.5 - 35.5	**Cian**	**5%**	$\theta_2 = \dfrac{360}{100}(5) = 3.6(5) = 18^\circ$
35.5 - 36.5	**Verde**	**7.5%**	$\theta_3 = \dfrac{360}{100}(7.5) = 3.6(7.5) = 27^\circ$
36.5 - 37.5	**Rosa**	**60%**	$\theta_4 = \dfrac{360}{100}(60) = 216^\circ$
37.5 - 38.5	**Naranja**	**25%**	$\theta_5 = \dfrac{360}{100}(25) = 3.6(25) = 90^\circ$
		100%	$\sum_{i=1}^{5} \theta_i = 360^\circ$



Para la construcción de este grafico se usó el applet en la dirección electrónica:http://www.eduteka.org/MI/master/interactivate/activities/Piechart/Index.html.

ACTIVIDAD 35

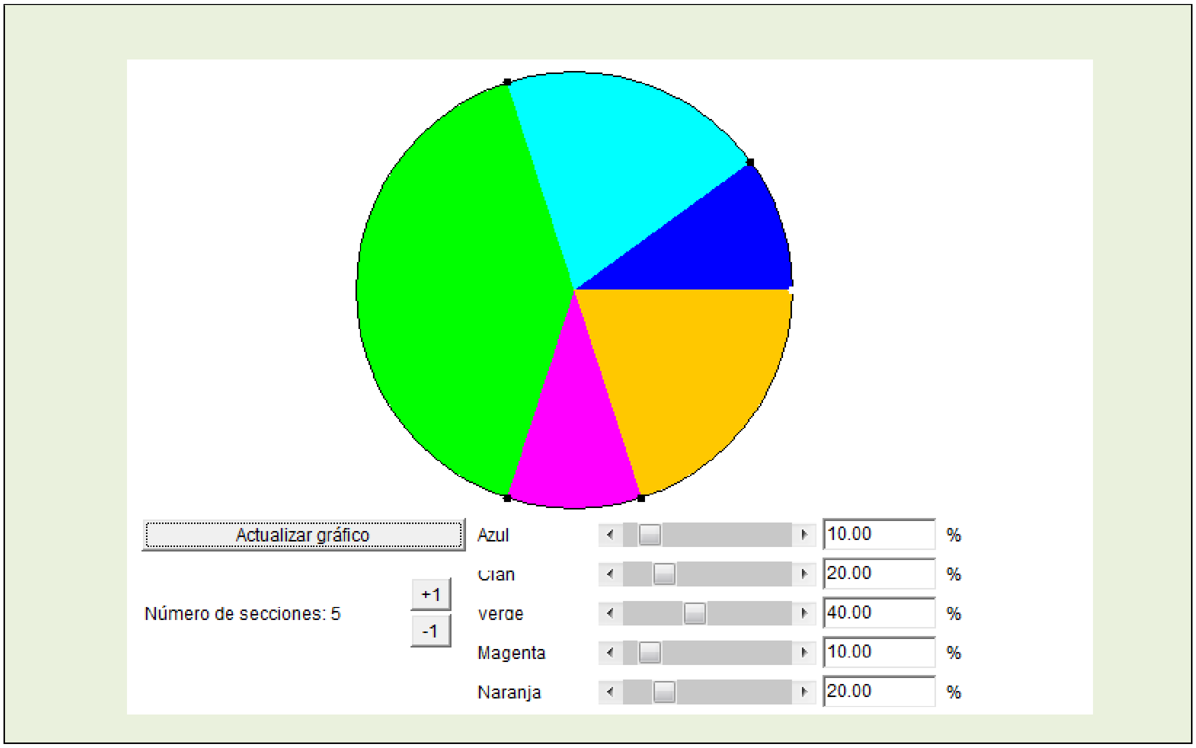
Construya el diagrama de pastel especificando los ángulos correspondientes. Considere para los datos en la siguiente tabla de frecuencias.

Categoría o Intervalo de Clase Real	Frecuencia absoluta f_i	Frecuencia acumulada porcentual
33.5 - 34.5	10	10%
34.5 - 35.5	20	20%
35.5 - 36.5	40	40%
36.5 - 37.5	10	10%

37.5 - 38.5	**20**	**20%**

Solución.

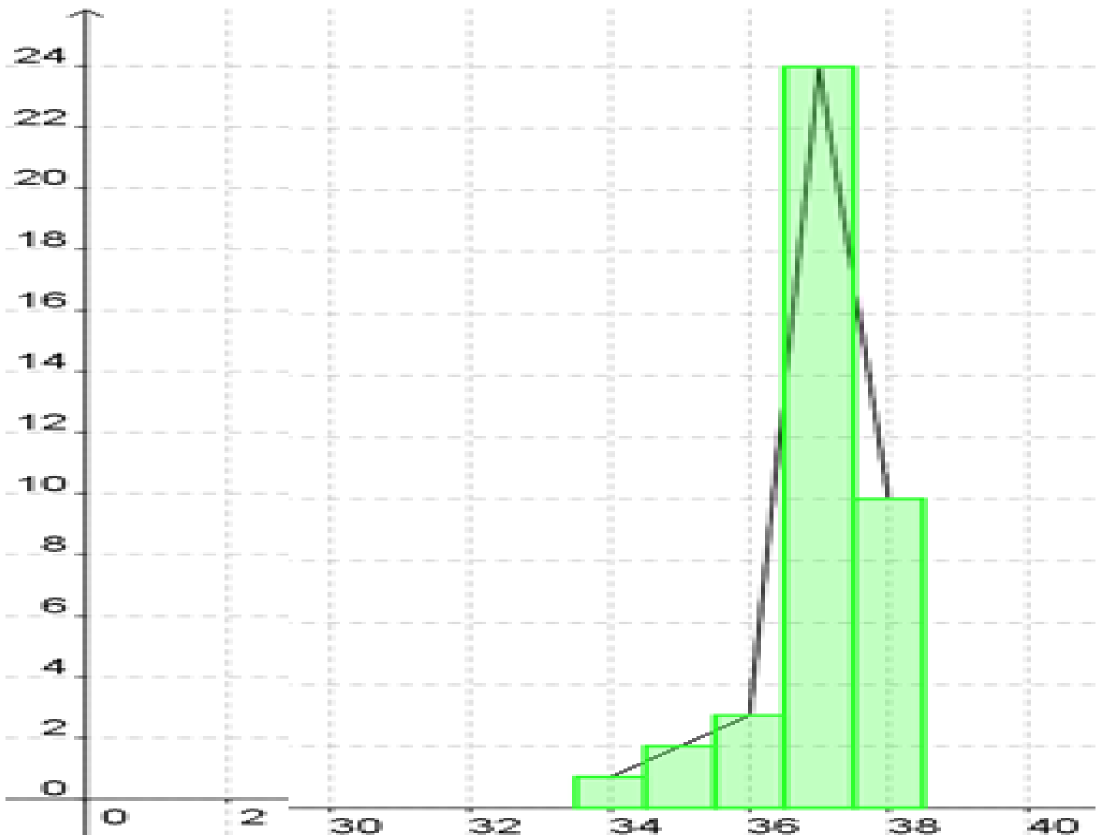
Categoría o Intervalo de Clase Real	**Color**	**frecuencia porcentual;** $f_{Ri}\%$,	**Angulo** $\theta = \frac{360}{100} f_{Ri}$
33.5 - 34.5	**Azul**	**10%**	$\theta_1 = \frac{360}{100}(10) = 3.6(10) = 36^{\circ}$
34.5 - 35.5	**Cian**	**20%**	$\theta_2 = \frac{360}{100}(5) = 3.6(20) = 72^{\circ}$
35.5 - 36.5	**Verde**	**40%**	$\theta_3 = \frac{360}{100}(40) = 3.6(40) = 144^{\circ}$
36.5 - 37.5	**Rosa**	**10%**	$\theta_4 = \frac{360}{100}(10) = 3.6(10) = 36^{\circ}$
37.5 - 38.5	**Naranja**	**20%**	$\theta_5 = \frac{360}{100}(20) = 3.6(20) = 72^{\circ}$



1.6.3 Polígono de frecuencias.

Para su construcción una vez que se dispone del diagrama de barras, localice la marca de clase en la base superior del rectángulo (barra), hágalo con todas las marcas de clase a continuación una con segmentos de recta todos los puntos.

A continuación se construirán esos gráficos con los datos de la tabla 1.12., Para la gráfica de barras, necesitamos las columnas; intervalos de clase real y frecuencias absolutas



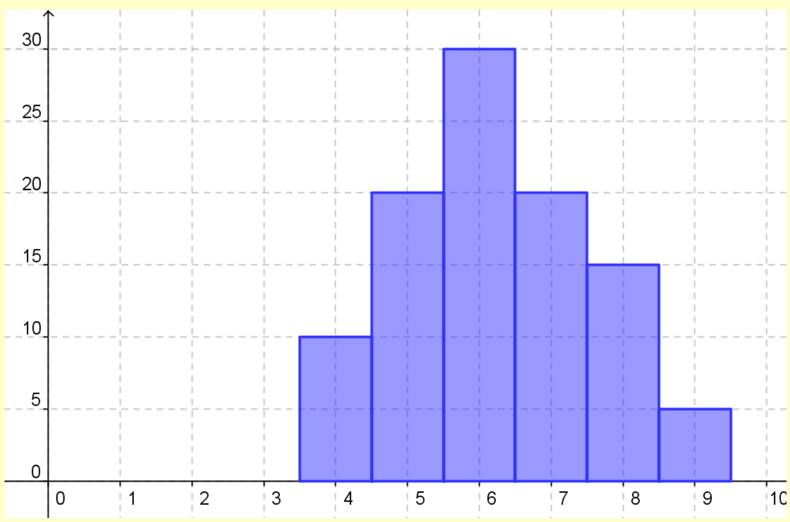
Estos diagramas pueden elaborarse con el paquete Geogebra, que puede ser bajado de manera gratuita en la página: http://www.geogebra.org/cms/

1.7.4 Curva de frecuencias

Para su construcción una vez que se dispone del diagrama de barras, localice la marca de clase en la base superior del rectángulo (barra), hágalo con todas las marcas de clase a continuación una con una curva suave y continua todos los puntos. Construya la curva de frecuencias para la siguiente tabla de datos agrupados.

Tabla 1.34 intervalos de clase real

Categoría o Intervalo de Clase Real	Frecuencia absoluta f_i
3.5 - 4.5	10
4.5 - 5.5	20
5.5 - 6.5	30
6.5 - 7.5	20
7.5 - 8.5	15
8.5 - 9.5	5



Trace una curva suave

1.7.5 Ojiva (creciente)

Se construye colocando las marcas de clase en el eje horizontal y en el eje vertical se representan las correspondiente frecuencias acumuladas relativas (en algunos casos se trabaja con las frecuencias absolutas acumuladas). También es posible construir la ojiva decreciente, pero por el momento es suficiente la ojiva creciente.

Solución.

ACTIVIDAD 36

Construya la ojiva decreciente, auxíliese con la tabla.

Categoría o Intervalo de Clase Real	**frecuencia absoluta** f_i	**frecuencia relativa** f_{Ri}	**frecuencia porcentual;** $f_{Ri}\%$,	**frecuencia acumulada absoluta** F_i,	**frecuencia acumulada relativa;** F_{Ri}	**frecuencia acumulada porcentual;** $F_{Ri}\%$	**marcas de clase;** m_i
33.5 - 34.5	**1**	**1/40**	**2.5%**	**1**	**1/40**	**2.5%**	**34**
34.5 - 35.5	**2**	**2/40**	**5%**	**3**	**3/40**	**7.5%**	**35**
35.5 - 36.5	**3**	**3/40**	**7.5%**	**6**	**6/40**	**15%**	**36**
36.5 - 37.5	**24**	**24/40**	**60%**	**30**	**30/40**	**75%**	**37**
37.5 - 38.5	**10**	**10/40**	**25%**	**40**	**40/40**	**100%**	**38**
	$n=40$		**100%**				

1.6.6 Diagrama de caja y bigote.

Se construye marcando primero los cuartiles Q1, Q2 y Q3, los cuales sirven para construir la caja, el inicio de la caja es Q1, el final de la caja es Q3, el bigote izquierdo es el segmento de recta con inicio en la variante menor de los datos, hasta Q1, el bigote derecho es un segmento de recta que va desde Q3 hasta la variante mayor de los datos.

En la tabla 1.12, correspondiente al problema 5, se obtuvieron columnas que aún no se han utilizado. Sin embargo son importantes en la construcción de las gráficas. Por lo pronto se retomara grafico de caja y bigotes visto en la sección 1.3.1

ACTIVIDAD 37

Considere nuevamente la tabla 1.12, construya un gráfico de caja y bigote.

Tabla 1.35 intervalos de clase real. Construcción de gráfica.

Categoría o Intervalo de Clase Real	**frecuencia absoluta** f_i	**frecuencia relativa** f_{Ri}	**frecuencia porcentual;** $f_{Ri}\%$,	**frecuencia acumulada absoluta** F_i,	***frecuencia acumulada relativa;*** F_{Ri}	**frecuencia acumulada porcentual;** $F_{Ri}\%$	**marcas de clase;** m_i
33.5 - 34.5	**1**	**1/40**	**2.5%**	**1**	***1/40***	**2.5%**	**34**
34.5 - 35.5	**2**	**2/40**	**5%**	**3**	***3/40***	**7.5%**	**35**
35.5 - 36.5	**3**	**3/40**	**7.5%**	**6**	***6/40***	**15%**	**36**
36.5 - 37.5	**24**	**24/40**	**60%**	**30**	***30/40***	**75%**	**37**
37.5 - 38.5	**10**	**10/40**	**25%**	**40**	***40/40***	**100%**	**38**
	$n=40$		**100%**				

Solución.

Para la construcción de la caja y bigote, se necesita determinar el valor máximo y mínimo y los cuartiles Q_1, Q_2, Q_3, el mínimo y el máximo son respectivamente: 33.5 y 38.5, la amplitud del intervalo es, $A_I = 1$, $n = 40$.
buscamos el cuartil 1 , a este le corresponde el 25 % de los datos (0.25) $n \cdot k = (40)\cdot(0.25) = 10$, buscamos el ordinal 10, cae en el cuarto Intervalo: **36.5 - 37.5**

$f_k = 24, F_{AAk} = 6, A_I = 1$, sustituyendo datos:

$$Q_1 = C_{0.25} = (36.5) + \left[\frac{(40)\cdot(0.25)-(6)}{(24)}\right](1) = 36.6667$$

buscamos el cuartil 2 , a este le corresponde el 50 % de los datos (0.50)

$n \cdot k = (40)\cdot(0.50) = 20$, buscamos el ordinal 20, cae en el cuarto Intervalo: **36.5 - 37.5**

$f_k = 24, F_{AAk} = 6, A_I = 1$, sustituyendo datos:

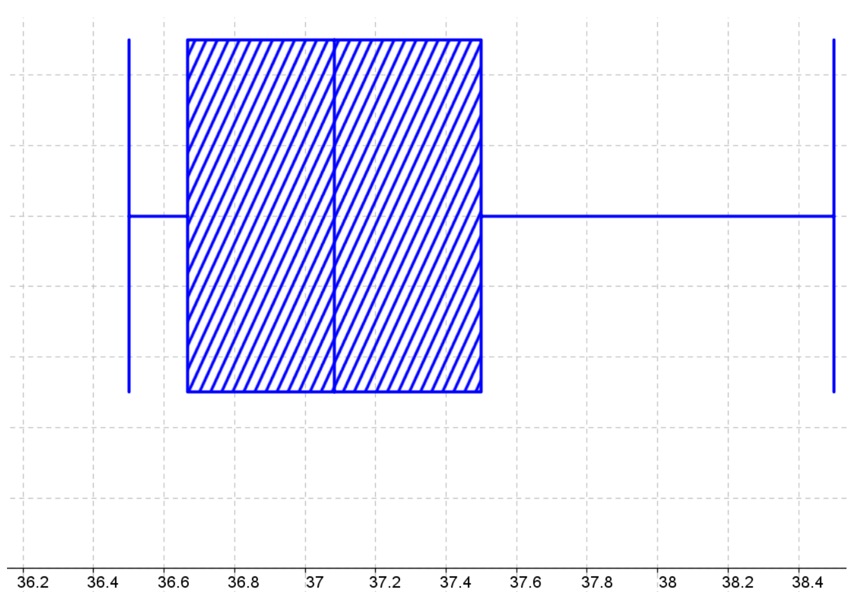
$$Q_2 = C_{0.50} = (36.5) + \left[\frac{(40)\cdot(0.50)-(6)}{(24)}\right](1) = 37.0833$$

buscamos el cuartil 3 , a este le corresponde el 75 % de los datos (0.75)

$n \cdot k = (40)\cdot(0.75) = 30$, buscamos el ordinal 30, cae en el cuarto Intervalo: **36.5 - 37.5**

$f_k = 24, F_{AAk} = 6, A_I = 1$, sustituyendo datos:

$$Q_3 = C_{0.50} = (36.5) + \left[\frac{(40)\cdot(0.75)-(6)}{(24)}\right](1) = 37.50$$



1.7 CONSTRUCCIÓN DE LA TABLA DE FRECUENCIAS PARA DATOS AGRUPADOS.

En los Actividades anteriores se ha establecido de antemano los intervalos o categorías, así como amplitudes de los mismos. Pero, ¿esto ha sido lo más adecuado?

Una regla general establece que el número de intervalos debe de estar entre cinco y veinte intervalos, lo cual complica la situación de elegir el mejor número de intervalos para tener la mejor presentación y agrupamiento de los datos. Un numero muy pequeño podría ocultar características o cualidades de los datos, lo mismo sucede si se elige un número muy grande de estos.

Existe una regla que permite elegir de manera óptima el número de intervalos, que permite observar características en la distribución de estos datos se lleva a cabo a través de la Regla de Sturges, $N_I = 1 + 3.33\log_{10} n$ aunque también existen otras reglas tales como la llamada regla de la raíz cuadrada.

En esta sección se presenta un algoritmo para la determinación del número de intervalos, determinando las amplitudes, de forma que los intervalos cumplan las características de ser exhaustivos y excluyentes

Tabla 1.36 Algoritmo para agrupación de datos.

ALGORITMO PARA CONSTRUCCION DE LA TABLA DE AGRUPACION DE DATOS.

1. Se calcula el rango: $R = V_M - V_m$,

2. se determinar el númerode intervalos con la regla de Sturges;

$N_I = 1 + 3.33\log_{10} n$

Para calcular calculamos;

3. la amplitud del intervalo es

4. El Rango nuevo podemos calcularlo como: $R_N = N_I \cdot A_I + (N_I - 1) \cdot U$, donde U, es la medida mínima de medición en los datos.

5. La diferencia de rangos entre dos: $\frac{\Delta R}{2} = \frac{R_N - R}{2}$

6. El límite inferior de la primera categoría o clase queda como: $L_{i1} = V_m - \frac{\Delta R}{2}$, mientras que el límite superior del primer intervalo
o categoría es: $L_{S1} = L_{i1} + A_I$. El proceso e hallar limites inferiores y superiores , se realiza tantas veces cono categorías se tengan.

7. El siguiente paso es la construcción de la Tabla de la distribución de frecuencias absolutas.

8. El siguiente paso es la construcción de la Tabla es agregar los intervalos de clase real, en la distribución de frecuencias absolutas. Para la primera categoría: $LR_{i1} = L_{i1} - \frac{U}{2}$

Para el límite superior del intervalo superior de la primera categoría: $L_{S1} = L_{i1} + \frac{U}{2}$, se hace este proceso tantas veces como categorías se tengan.

9. también será necesario calcular otras cantidades que serán necesarias para realizar otros cálculos, para datos agrupados; las medidas centrales y de dispersión. Para este propósito es necesario determinar las marcas de clase, la frecuencia relativa, la frecuencia acumulada, la frecuencia relativa acumulada, la cuales pueden calcularse a través de:

10. Finalmente tenemos dentro del algoritmo, se construyen las representaciones gráficas: Gráfica de barras, Polígonos de frecuencias, Ojivas, Curvas de frecuencias, Histogramas, Gráfica circular, Gráfica de caja y bigote

PROBLEMA 6

La siguiente información se obtuvo después de medir los rendimientos en el consumo de gasolina en automóviles ATHOS modelo 2010, las mediciones están en kilómetros por litro de combustible y fueron obtenidos en una muestra de 40 automóviles.

Tabla 1.37 de consumo de gasolina km por litro

12.70	14.13	15.02	16.04
12.70	14.23	15.52	16.44
12.81	14.53	15.62	16.44
13.11	14.53	15.73	16.52
13.11	14.54	15.83	16.61
13.21	14.54	15.83	16.72
13.31	14.84	15.83	17.05
13.42	14.84	15.83	17.17
13.52	14.94	15.83	17.29
13.72	15.02	16.04	17.58

La agrupación de datos tiene por objetivo determinar las características de la población. Tales como puntos máximos, mínimos, donde es creciente o decreciente, la forma que tiene etc. Según otros autores, se logra el objetivo de resumir los datos.

Dada la tabla de la distribución de frecuencias, es posible hallar medidas de tendencia central, así como también medidas de dispersión.

Considera los siguientes datos, organízalos, y agrupados en categorías, de acuerdo a lo indicado en la sección 1.2.3 (tablas de distribución de frecuencias).

La siguiente información se obtuvo después de medir los rendimientos en el consumo de gasolina en automóviles ATHOS modelo 2010, las mediciones están en kilómetros por litro de combustible y fueron obtenidos en una muestra de 40 automóviles.

1. Determine las medidas de tendencia central para:	2. Determine las medidas de Dispersión absoluta y relativa para:
a. datos no agrupados.	b. datos no agrupados.
c. Datos agrupados	d. Datos agrupados

e.. compare ambos resultados.

Considerando los datos, previamente ordenados, se puede observar que el rango de la variable es el rendimiento en kilómetros de un litro de gasolina en Automóviles ATHOS. A continuación se muestran los pasos del algoritmo para la agrupación de datos.

Aplicación del algoritmo para construcción de la tabla de datos agrupados.

1. $R = V_M - V_m = 17.58 - 12.70 = 4.88$, nuestro siguiente paso es determinar el número de intervalos con la regla de Sturges;

2. $N_I = 1 + 3.33\log_{10} n$; se dispone de 40 datos, lo cual nos da un valor de 6.3348, y que redondeamos a 7.

Para calcular la amplitud del intervalo calculamos;

3. $A_I = \dfrac{R}{N_I} = \dfrac{4.88}{7} = 0.6971 \approx 0.70$, la justificación de este redondeo, es porque la información es del orden de centésimas.

4. El Rango nuevo podemos calcularlo como:
$R_N = N_I \cdot A_I + (N_I - 1)\cdot U = 7\cdot(0.70) + 6\cdot(0.01) = 4.96$, donde $U = 0.01$, es la medida mínima de medición en los datos.

5. La diferencia de rangos entre dos: $\dfrac{\Delta R}{2} = \dfrac{R_N - R}{2} = \dfrac{4.96 - 4.88}{2} = \dfrac{0.08}{2} = 0.04$

6. El límite inferior de la primera categoría queda como: $L_{i1} = V_m - \dfrac{\Delta R}{2} = 12.70 - 0.04 = 12.66$, mientras que el límite superior del primer intervalo o categoría es: $L_{S1} = L_{i1} + A_I = 12.66 + 0.70 = 13.36$

7. El siguiente paso es la construcción de la Tabla de la distribución de frecuencias absolutas. Completando lo faltante:

Tabla 1.38 de distribución de frecuencias. Datos agrupados.

Intervalo de Clase I.C.	Frecuencia (Absoluta) f_i
12.66 – 13.36	7
13.37 - 14.07	3
14.08 - 14.78	6
14.79 _ 15.49	5
15.50 _ 16.20	10
16.21 _ 16.91	5
16.92 - 17.62	4
	$n = 40$

8. El siguiente paso es la construcción de la Tabla es agregar los intervalos de clase real, en la distribución de frecuencias absolutas.

Tabla 1.39 de distribución de frecuencias. Datos agrupados.

Intervalo de Clase I.C.	Intervalo de Clase I.C.R.	Frecuencia (Absoluta) f_i
12.66 – 13.36	12.655 – 13.365	7
13.37 - 14.07	13.365 - 14.075	3
14.08 - 14.78	14.075 - 14.785	6
14.79 _ 15.49	14.785 _ 15.495	5
15.50 _ 16.20	15.495 _ 16.205	10
16.21 _ 16.91	16.205 _ 16.915	5
16.92 - 17.62	16.915 - 17.625	4
		$n = 40$

9. también será necesario calcular otras cantidades que serán necesarias para realizar otros cálculos, para datos agrupados; las medidas centrales y de dispersión. Para este propósito es necesario determinar las marcas de clase, la frecuencia relativa, la frecuencia acumulada, la frecuencia relativa acumulada,la cuales pueden calcularse a través dela tabla e frecuencias.

ACTIVIDAD 38

Considera los daos de la tabla 1.16, y completa lo que se indica en la siguiente tabla.

Intervalo de Clase	I.C.
Intervalo de Real	I.C.R.
Frecuencia (Absoluta)	f_i
Frecuencia (Relativa)	f_{Ri}
Frecuencia Porcentual	f_{Ri} %
Frecuencia Acumulada Absoluta	F_i
Frecuencia Acumulada Relativa	F_i

Solución.

I.C.	I.C.R.	m_i	f_i	f_{Ri}	f_{Ri} %	F_i	F_i
12.66 – 13.36	12.655– 13.365		7	$\frac{7}{40}$	17.5%	7	$\frac{7}{40}$
13.37- 14.07	13.365 - 14.075	13.74		$\frac{3}{40}$			
14.08 - 14.78	14.075 - 14.785		6		15%	16	

14.79 _ 15.49	14.785_ 15.495	15.20		$\frac{6}{40}$			
15.50 _ 16.20	15.495 _ 16.205			$\frac{9}{40}$	22.5%	31	
16.21 _ 16.91	16.205_ 16.915		5				$\frac{36}{40}$
16.92 - 17.62	16.915 - 17.625	17.39		$\frac{4}{40}$	10%		$\frac{40}{40}=1$
			$n=\sum_{i=1}^{n} f_i$ $=40$	$\sum_{i=1}^{n} f_{Ri}$ $=1$	100%		

Algunas personas consideran erróneamente que la agrupación de datos tiene por finalidad reducir un poco la cantidad de información, y de esta manera hacer más manejable la cantidad de datos.

Sin embargo la razón por la cual se realiza el trabajo de agrupar los datos consiste en que en este proceso de compactación de datos presenta ventajas inmediatas tales como observación de características con mayor facilidad, tales como la forma de la curva de frecuencias, por ejemplo, si esta es unimodal, multimodal, el tipo de sesgamiento; positivo o negativo y la curtosis (grado de agudeza), o por si por el contrario tiene varios máximos o tiene otra forma; tipo u o jota o jota invertida y en particular la localización de valores máximos o mínimos, tendencias en el crecimiento o decrecimiento.

ACTIVIDAD 39

Para los siguientes conjuntos de datos, construya la tabla de frecuencias.

Solución.

a.

1	3	5	8	11	14	18	22
1	3	6	9	11	14	18	23
2	3	7	9	12	15	19	24
2	3	8	10	13	16	20	25
2	4	8	10	13	17	21	26

b.

5.1	5.3	5.4	6.0	6.1	6.3	6.6	6.6
5.2	5.3	5.4	6.1	6.1	6.4	6.6	6.6
5.3	5.3	5.5	6.1	6.1	6.5	6.6	6.6
5.3	5.3	5.5	6.1	6.2	6.5	6.6	6.7
5.3	5.4	5.6	6.1	6.3	6.6	6.6	6.8

c.

12.70	12.70	12.70	12.70	12.70	12.70
12.70	12.70	12.70	12.70	12.70	12.70
12.81	12.81	12.81	12.81	12.81	12.81
13.11	13.11	13.11	13.11	13.11	13.11
13.11	13.11	13.11	13.11	13.11	13.11

d.

0.001	0.004	0.006	0.015	0.033
0.002	0.004	0.006	0.015	0.033
0.003	0.004	0.008	0.015	0.033
0.003	0.005	0.008	0.015	0.033
0.003	0.006	0.008	0.023	0.033
0.003	0.006	0.008	0.023	0.033
0.003	0.006	0.008	0.023	0.045
0.003	0.006	0.008	0.023	0.045
0.003	0.006	0.008	0.023	0.055
0.003	0.006	0.008	0.032	0.060

Por otra parte una desventaja que se presenta como consecuencia del proceso, es la perdida de precisión en el cálculo de las medidas centrales y de dispersión, como podremos constatar más adelante.

ACTIVIDAD 40

Para los datos de la tabla 1.16, determina:

Solución.

media	coeficiente de variación porcentual
mediana	cuartil 1
moda	cuartil 2
Media armónica	cuartil 3
media geométrica	decil 2
varianza	decil 7
Desviación estándar	porcentil 86
coeficiente de variación	Porcentil 17
rango	Rango semicuartilico

En estadística descriptiva, se denomina rango intercuartílico o rango intercuartil, a la diferencia entre el tercer y el primer cuartil.

$$R_Q = Q_3 - Q_1$$

Finalmente tenemos dentro del algoritmo:
10. Representaciones gráficas: Gráfica de barras, Polígonos de frecuencias, Ojivas, Curvas de frecuencias, Histogramas, Gráfica circular. Gráfica de caja y bigote.

La construcción de la gráfica es necesaria para observar características o detalles referentes a como se distribuyen los datos.

Construye la gráfica de columnas (o barras) considerado las siguientes indicaciones: en el eje horizontal representarás los intervalos de clase y en el eje vertical la frecuencia (puede ser la absoluta o la frecuencial), en cualquiera de los casos deberás indicar de cual se trata la gráfica de barras de frecuencias absolutas o la gráfica de barras de frecuencias relativas.

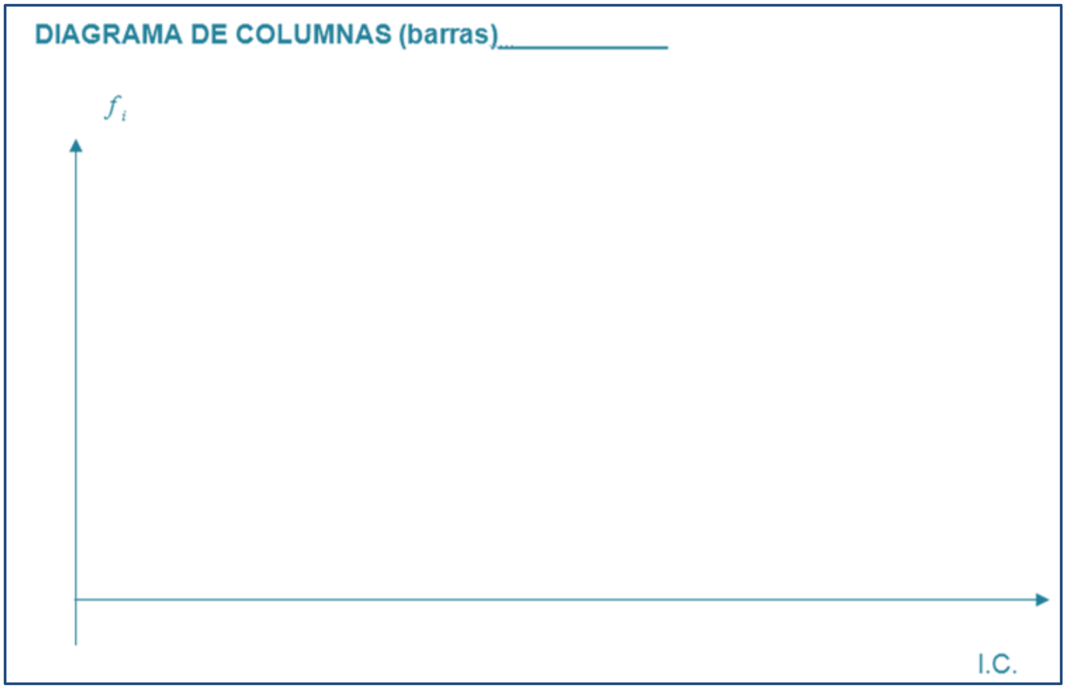
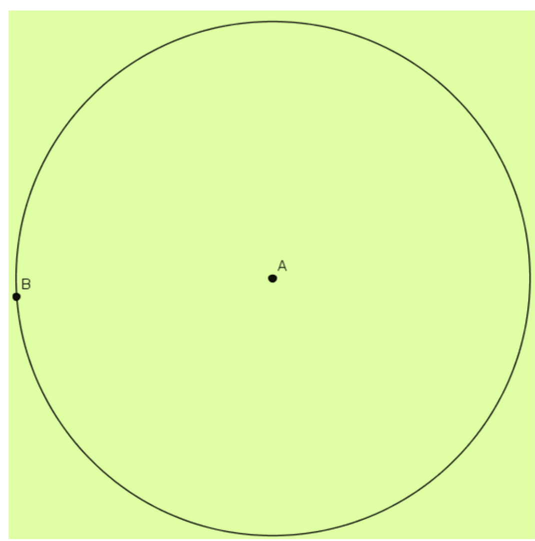


DIAGRAMA CIRCULAR

FIGURA 2 DIAGRAMA DE PASTEL (DIAGRAMA DE PAY)

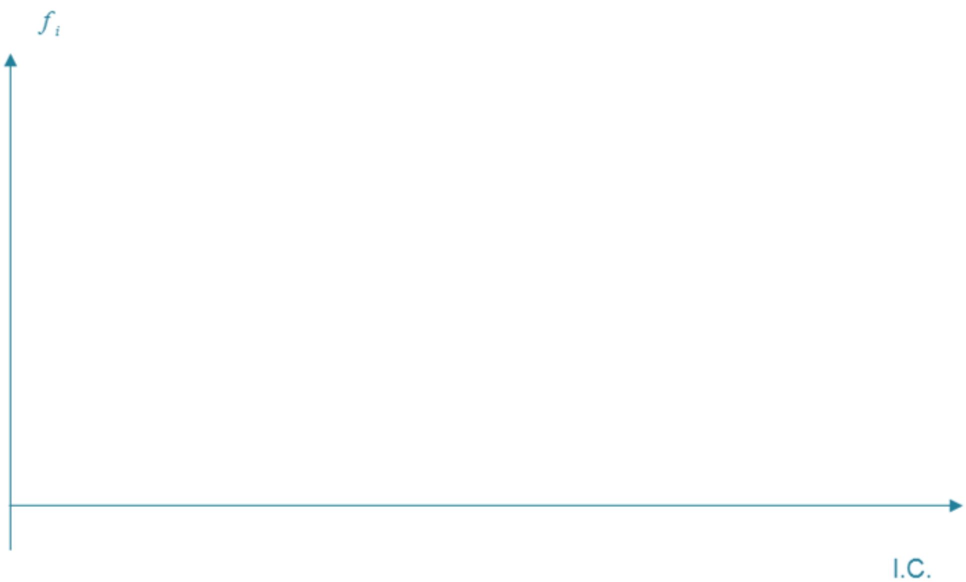
Construye un diagrama de pastel, para los siete intervalos, cada porción deberá ser proporcional a la cantidad de datos de cada intervalo de clase.



CONSTRUCCION DEL POLÍGONO DE FRECUENCIAS Y DE LA CURVA DE FRECUENCIAS.

Con la información obtenida en la tabla de datos estadísticos agrupados, refleja el valor de la marca de clase sobre la base superior de cada rectángulo, Une los puntos por medio de segmentos de línea recta. Si en vez de segmentos se le ajusta una curva, la gráfica recibe el nombre de curva de frecuencias.

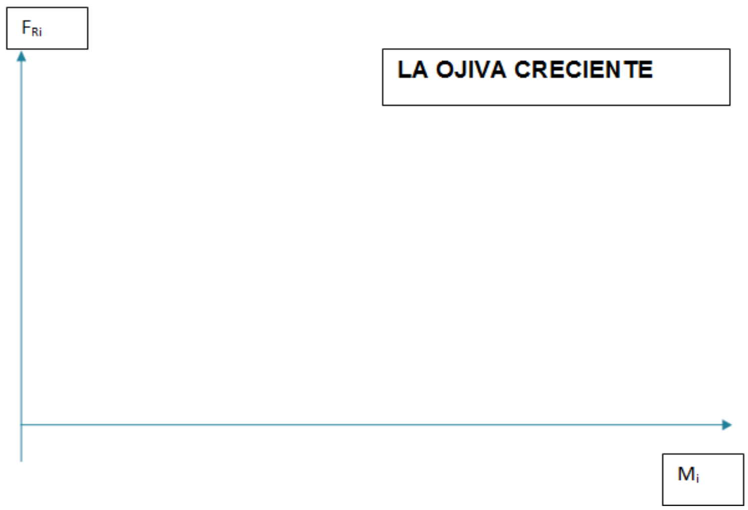
POLÍGONO DE FRECUENCIAS



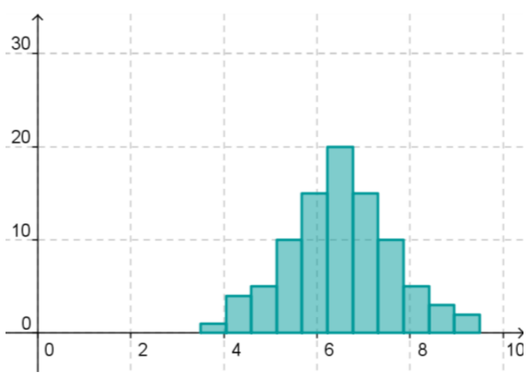
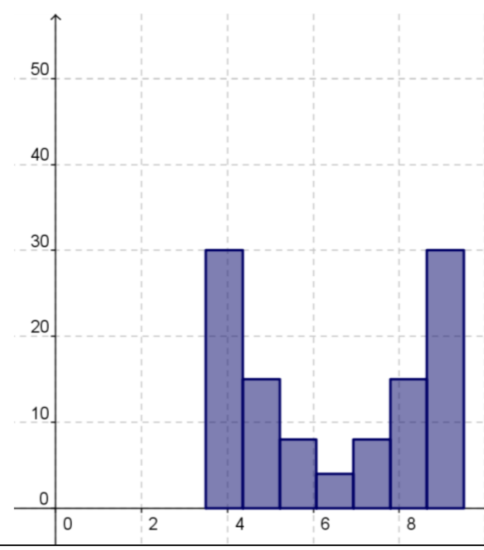
GRAFICA DE LA CURVA DE FRECUENCIAS

OJIVA CRECIENTE

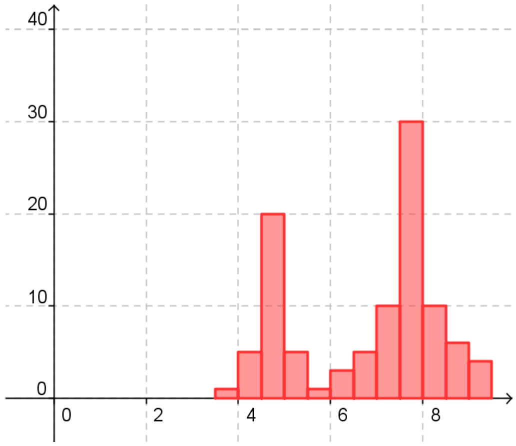
También es posible construir un gráfico que te permita obtener la mediana, esta recibe el nombre de ojiva, creciente o decreciente. Para construir la ojiva creciente es necesario considerar los datos en las columnas marcadas como F_{ri} frecuencia relativa acumulada y $m_{i,}$, la marca de clase . Los datos de F_{ri} e les representa en el eje vertical, mientras que los de m_i se les representa en el eje horizontal.Construye la ojiva creciente.



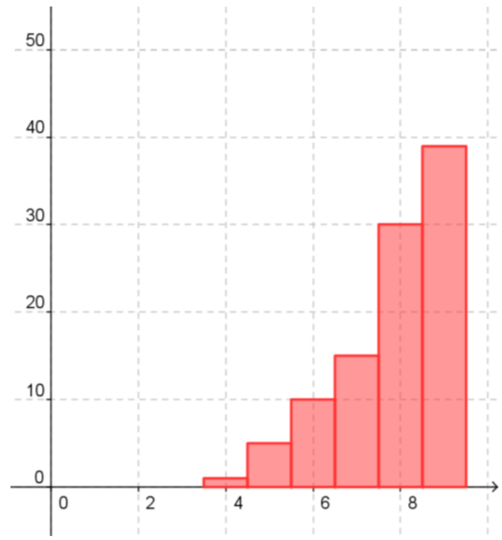
1.7.1 TIPOS DE DISTRIBUCIONES DE DATOS.

Unimodal	FORMA DE U
	

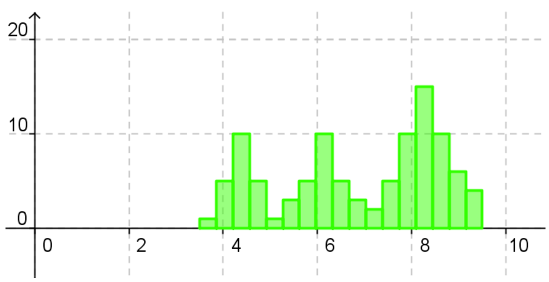
BIMODAL



FORMA DE J



MULTIMODAL



FORMA DE J INVERTIDA