REPÚBLICA DE COLOMBIA
SECRETARÍA DE EDUCACIÓN MUNICIPAL DE PALMIRA
“INSTITUCIÓN EDUCATIVA “DE ROZO”
Aprobada por Resolución Nº 0835 del 20 de FEBRERO de 2.017

INFORMACIÓN GENERAL GUÍA DE APRENDIZAJE No. 1	
ÁREA O ASIGNATURA:	**Estadística**
NOMBRE DE LA GUIA(S):	Medidas de tendencia central para datos agrupados
DURACIÓN (MES)	12 marzo – 12 de Abril
DOCENTE(S):	Duivan Anderson Alvarez - Héctor Fabio Buitrago
GRADO:	**ONCE**
PERIODO:	Uno
OBJETIVO DE APRENDIZAJE y/o DBA:	**DBA 10** Plantea y resuelve problemas en los que se reconoce cuando dos eventos son o no independientes y usa la probabilidad condicional para comprobarlo. **PENSAMIENTO ALEATORIO, ESTANDAR 4** Describo tendencias que se observan en conjuntos de variables relacionadas.

INTRODUCCIÓN: Leer cuidadosamente toda la introducción

Esta guía tendrá como objetivo comprender reafirmar conocimientos de la medida de tendencia central y medidas de dispersión ya que en el diagnostico hecho a los estudiantes arrojo una debilidad de más del 50%.
La guía contiene una parte en la cual se dan las explicaciones, y se encuentra en las secciones que dice: **¿Qué voy a aprender?** Y **Lo que estoy aprendiendo**. Luego viene una sección en la que se hacen las actividades y esa sección se llama **practico lo que aprendí** en donde estarán las actividades por realizar, la última sección es una autoevaluación para dar sus impresiones y evaluar su propio aprendizaje, así como la guía misma esta sección se llama **¿cómo sé que aprendí?** y **que aprendí**.

Que deberia saber

Resuelva los siguientes problemas:

Problema 1

Un estudiante de primer semestre de universidad pregunta a 10 compañeros sobre la cantidad de dinero que gastan a diario. Los resultados fueron:

5.000, 7.500, 10.000, 8.000, 5.000, 10.000, 8.000, 50.000, 5.000, 6.500

Caracterizar la variable usando las medidas de tendencia central.

Problema 2

El departamento de control de calidad de una embotelladora de refrescos quiere evaluar dos de sus máquinas de llenado. Para ello, se mide el proceso de llenado de 20 refrescos en cada una de las máquinas. Los resultados se muestran en las tablas presentadas al ladillo.

Si se ofrece al público refrescos con un contenido de 350 mililitros, determinar:

a. ¿Cuál de las dos máquinas cumple mejor, en promedio, con este requisito?

b. ¿Cuál de las máquinas tiene mayor regularidad en el llenado? Utilizar la desviación estándar para justificar la respuesta.

Trate de resolver cada uno de los anteriores problemas hasta donde más pueda y en caso que ya no pueda, encontrara la solución de estos problemas en la parte final de la guía que dice anexo.

¿Qué voy a aprender?

A resolver problemas que impliquen medidas de tendencia central, así como problemas relacionados con medidas de dispersión.

Lo que estoy aprendiendo

Media aritmética o promedio

La **media aritmética** en un conjunto de datos agrupados por clases es el cociente entre la suma de todos los productos de las marcas de clase o puntos medios de cada intervalo (C_k) por la frecuencia absoluta (f_k) correspondiente, y el total de datos (N).

La media aritmética viene dada por la siguiente expresión:

$$\bar{x} = \frac{\sum_{k=1}^{m} (C_k \cdot f_k)}{N}$$

Mediana

Para calcular la **mediana** para datos agrupados por clases es necesario ubicar primero en la distribución de frecuencias el intervalo en donde se encuentra la mediana. La manera de calcularla es encontrando la posición $\frac{N}{2}$. Este intervalo se conoce como **intervalo de la mediana**.

La mediana de un conjunto de datos agrupados se calcula así:

$$Me = L_k + \frac{\left(\frac{N}{2} - F_{k-1}\right)}{f_k} \cdot A$$

L_k es el límite inferior del intervalo mediano, N es el tamaño de la muestra, F_{k-1} es la frecuencia acumulada anterior al intervalo mediano, f_k es la frecuencia absoluta del intervalo mediano y A es la amplitud del intervalo de la mediana.

Moda

La **moda** de una variable estadística es el valor de la variable que presenta mayor frecuencia absoluta. La moda se representa por *Mo*.

Si los datos aparecen en clases, se toma como valor aproximado de la moda la marca de clase de la **clase modal**.

Problema de aplicación solucionado

En la tabla se muestra el peso en kilogramos de los 61 deportistas de una liga del Valle.

Peso (kg)	Marcas de clase (C_i)	f_i	F_i
[50, 60)	55	5	5
[60, 70)	65	12	17
[70, 80)	75	13	30
[80, 90)	85	17	47
[90, 100)	95	10	57
[100, 110)	105	4	61

Tabla 6.6

Calcular el peso promedio del deportista, el peso más frecuente de los deportistas y el peso central de los deportistas.

Solución

El peso promedio se halla con la media aritmética

$$\bar{x} = \frac{55x5 + 65x12 + 75x13 + 85x17 + 95x10 + 105x4}{61} =$$

$$= \frac{4845}{61} = 79,42$$

En este problema en el cuadro la marca de clase "es el punto medio" de cada Intervalo, se representa con C_i. El intervalo en este problema es el peso (Kg) (50,60) la marca de clase seria 55 como se puede observar en la tabla. La f_i es la frecuencia absoluta de cada marca de clase. Puede observarse que se explica mediante unas líneas la relación de la tabla y la operación para que entiendan el procedimiento.

Explicación del resultado obtenido.

Si los 61 deportistas tuvieran el mismo peso este sería de 79,42

Cálculo de la mediana "Dato central"

Para calcular la mediana procederemos de la siguiente manera leer muy bien cada paso que se realizara a continuación:

Explicación de la tabla relacionado con las variables de la formula de la mediana

Frecuencia absoluta del intervalo mediano

frecuencia absoluta acumulada F_i

Peso (kg)	Marcas de clase (C_i)	f_i	F_i
[50, 60)	55	5	5
[60, 70)	65	12	17
[70, 80)	75	13	30
[80, 90)	85	17	47
[90, 100)	95	10	57
[100, 110)	105	4	61

Frecuencia acumulada anterior al intervalo mediano

Intervalo de la mediana

Numero de datos total
N=61

En la tabla el intervalo [80,90) resaltado es el intervalo de la mediana, este se calcula dividiendo el número de datos totales entre 2, es decir 61 dividido entre 2 y daria 30,5, esta cantidad la miramos en la columna de las frecuencias acumuladas y lo mas cercano es 30 por tanto estara por debajo de este número es decir en la fila donde esta el 47, alli estara ubicado el intervalo de la mediana[80,90) donde 80 es el limite inferior y 90 es el limite superior

Formula de la mediana con los nombres de las variables, para entender mejor el procedimiento es conveniente mirar la fórmula con lo anteriormente expuesto

Número de datos

Frecuencia absouta acumulada anteror al intervalo mediano

$$Me = L_k + \frac{\left(\frac{N}{2} - F_{k-1}\right)}{f_k} \cdot A$$

Amplitud del intervalo

Limite inferior

Frecuancia absoluta del intervalo mediano

Solución

Entonces, $Me = 80 + \frac{\left(\frac{61}{2} - 30\right)}{17} \cdot 10 = 80{,}29$

Esto significa que el 50% de los deportistas de la liga tienen un peso por encima de 80,3 kg y el otro 50% por debajo de tal valor.

- En esta situación el intervalo modal coincide con el intervalo de la mediana; por lo tanto: $Mo = \frac{80 + 90}{2} = 85.$ puesto que el 80,29 está en el intervalo de 80 a 90

Medidas de dispersión

Actividad de introducción

Una finca cafetera de Nariño recolecta durante 5 días de la semana 15 kg, 18 kg, 21 kg, 23 kg y

13 kg de grano, mientras que otra del Cauca recolecta en el mismo lapso de tiempo 8 kg, 12 kg, 26 kg, 30kg y 5 kg. ¿Cuáles de estos datos se encuentran más dispersos, los de la producción de la finca de Nariño o los de la finca del Cauca?

Trata de responder la actividad anterior y explica tu respuesta.

Vamos a analizar el siguiente problema que nos introducirá al objeto de aprendizaje que estudiaremos a continuación.

En la siguiente tabla se muestran las distancias (en metros), que alcanzan dos atletas de salto largo en un día de entrenamiento.

Atleta 1	Atleta 2
12,3	14,2
14,5	13,9
11,7	12,8
15,1	12,7
12,8	14,1
11,9	14
15,3	12,9
13,7	13,7
13,9	13,1
12,8	12,6

Tabla 6.12

¿Cómo determinarías cual es el atleta mas regular o constante en sus saltos?

Antes de mirar la solución reflexiona y trata de dar una respuesta.

Solución:

La media o promedio de la distancia alcanzada por cada atleta es:

Atleta 1

$$\bar{x} = \frac{134}{10} = 13{,}4$$

Atleta 2

$$\bar{x} = \frac{134}{10} = 13{,}4$$

Como el promedio de ambos atletas es el mismo, esta medida no permite comparar los resultados. Una alternativa para determinar cuál de los dos es más regular, es observar la **variabilidad** o la **dispersión** en las distancias alcanzadas por cada uno. Para ello, se organizan de menor a mayor los registros de ambos:

Atleta 1	11,7	11,9	12,3	12,8	12,8	13,7	13,9	14,5	15,1	15,3
Atleta 2	12,6	12,7	12,8	12,9	13,1	13,7	13,9	14	14,1	14,2

Tabla 6.13

Se puede advertir que el Atleta 1 es quien más varía en las distancias que alcanza. Por lo tanto, el Atleta 2 es más constante en sus saltos.

Las medidas de dispersión muestran la variabilidad de una distribución indicando cuán alejados están los datos de la media. Cuanto mayor sea ese valor, mayor será la variabilidad, y cuanto menor sea, más homogénea será.

Rango

El rango o recorrido de una distribución de datos es la diferencia entre el valor máximo ($x_{máx}$) y el valor mínimo ($x_{mín}$).

$$\text{Rango} = x_{máx} - x_{mín}$$

Ejemplo

En la tabla 6.14, se muestran las estaturas de un grupo de estudiantes compuesto por 24 mujeres. El rango corresponde a la diferencia entre 168 (valor máximo) y 149 (valor mínimo).

Es decir, rango = 168 − 149 = 19.

Estaturas (cm)	No. de mujeres
[149, 153)	2
[153, 157)	3
[157, 161)	7
[161, 165)	8
[165, 168)	4

Tabla 6.14

Desviación respecto a la media

La **desviación respecto a la media**, denotada d_k, corresponde a la diferencia entre cada valor de la variable x_k y la media aritmética $\overline{x}$.

$$d_k = x_k - \overline{x}$$

En datos agrupados por clases es la diferencia entre la marca de clase de cada intervalo y la media aritmética $\overline{x}$.

Esta medida da información de lo cerca o lejos que está un dato de los demás datos del conjunto. El signo de esta medida indica si el valor está por encima de la media (signo positivo) o por debajo de la media (signo negativo).

Desviación media

La **desviación media** para datos agrupados por clases, denotada D_k, corresponde a la media de las desviaciones con respecto a la media y se calcula con la fórmula:

$$D_k = \frac{\sum_{k=1}^{m} \left| C_k - \overline{x} \right| \cdot f_k}{N}$$

Por lo tanto, para determinar la desviación media para datos agrupados por clases, se deben calcular las marcas de clase de cada intervalo y luego se multiplican por las frecuencias absolutas, estos datos vienen agrupados en una **tabla de frecuencias** como se vera en un ejemplo despues.

Varianza

La **varianza** para datos agrupados por clases, denotada s^2, es la media aritmética de cuadrados de las desviaciones con respecto a la media. Se halla mediante la expresión:

$$s^2 = \frac{\sum_{k=1}^{m} \left(C_k - \bar{x}\right)^2 \cdot f_k}{N}$$

La varianza permite identificar la diferencia media que hay entre cada uno de los valores respecto a la media del conjunto de datos.

Desviación típica

La **desviación típica** para datos agrupados por clases, s, es la raíz cuadrada positiva de la varianza y se halla a través de la siguiente expresión:

$$s = \sqrt{\frac{\sum_{k=1}^{m} \left(C_k - \bar{x}\right)^2 \cdot f_k}{N}}$$

Ejemplo

Los datos de la Tabla 6.15 corresponden a las temperaturas diarias registradas a las 2:00 p.m. durante un mes en la ciudad de Manizales.

Temperatura (ºC)	Frecuencia absoluta f_k	Marcas de clase C_k
[16; 18,8)	5	17,4
[18,8; 21,6)	7	20,2
[21,6; 24,4)	14	23
[24,4; 27,2)	3	25,8
[27,2; 30)	2	28,6
Sumatoria		**685**

Tabla 6.15

La media aritmética de estos datos corresponde a $\overline{x} = \frac{685}{31} = 22{,}1$.
Con este dato, se completa la Tabla 6.16.

Temperatura (ºC)	f_k	C_k	d_k	d_k^2	$d_k^2 \cdot f_k$
[16; 18,8)	5	17,4	-4,7	22,09	110,45
[18,8; 21,6)	7	20,2	-1,9	3,61	25,27
[21,6; 24,4)	14	23	0,9	0,81	11,34
[24,4; 27,2)	3	25,8	3,7	13,69	41,07
[27,2; 30)	2	28,6	6,5	42,25	84,5

Tabla 6.16

Se halla la varianza s^2 y, posteriormente, la desviación típica s:

$$s^2 = \frac{110{,}45 + 25{,}27 + 11{,}34 + 41{,}07 + 84{,}5}{31} = \frac{272{,}63}{31} = 8{,}79.$$

$$s = \sqrt{8{,}79} = 2{,}96$$

Cuanto mayor sean la varianza y la desviación típica, más dispersos estarán los datos.

Práctico lo que aprendí

Resolución de problemas

1 En una fábrica se evalúa el tiempo de duración de un nuevo tipo de bombilla LED, para ello se analizaron 100 bombillas (Tabla 6.7).

Tiempo (horas)	Número de bombillas
[1000, 3500)	14
[3500, 5000)	25
[5000, 7500)	31
[10000, 12500)	18
[12500, 15000)	12

Tabla 6.7

3 Dada la distribución estadística definida por la tabla 6.18.

x_i	[0, 5)	[5, 10)	[10, 15)	[15, 20)	[20, 25)	[25, 30)
f_i	5	6	8	11	1	13

Tabla 6.18

a. Calcula la media, la mediana y la moda.

b. Halla la varianza y la desviación típica.

2 La tabla muestra las estaturas de 40 funcionarios de una empresa.

Estatura (m)	Número de funcionarios
[1,46; 1,53)	4
[1,53; 1,60)	9
[1,60; 1,67)	10
[1,67; 1,74)	8
[1,74; 1,81)	9

Tabla 6.17

a. Determina el rango del conjunto de datos.

b. Calcula la desviación con respecto a la media de cada intervalo y escribe cuál está más alejado de la media.

c. Halla la varianza y la desviación típica.

4 Se dan dos conjuntos de datos.

A: 1, 3, 5, 7, 9

B: 1, 5, 10, 15, 30

Sin necesidad de hacer ningún cálculo, ¿cuál de los dos conjuntos tiene mayor dispersión?

¿Cómo sé que aprendí?

1 Las tablas 6.19 y 6.20 muestran los resultados de dos laboratorios cuando analizan la cantidad de residuos secos en el agua potable (mg/L).

Laboratorio A

Residuos secos	f_k
[8; 10)	15
[10; 12)	8
[12; 14)	7
[16; 18)	6
[18; 20)	24

Tabla 6.19

2 A continuación se muestran las edades de motociclistas cuando fallecieron en accidentes de tránsito.

Edad (años)	Número de motociclistas
[15, 21)	101
[21, 27)	253
[27, 33)	137
[33, 39)	211
[39, 45)	116

Tabla 6.8

Escribe falso o verdadero:

a. El 50% de los motociclistas que fallecieron tenían edades por debajo de los 31 años.

b. La edad promedio de los motociclistas que fallecieron es 31,03 años.

c. La moda se encuentra entre 33 y 39 años.

Laboratorio B

Residuos secos	f_k
[8; 10)	30
[10; 12)	8
[12; 14)	11
[16; 18)	4
[18; 20)	4

Tabla 6.20

Se elegirá el laboratorio que tenga resultados que presenten menos variabilidad en sus datos. ¿Qué decisión consideras que es la mas adecuada? Justifica tus respuestas.

¿Qué aprendí?

Vas a reflexionar respecto a cómo te sentiste y qué tanto aprendiste en el desarrollo de esta guía.

En tu cuaderno registra las conclusiones a las que llegaste *¡Debes de ser muy sincero!*

1. ¿Qué fue lo que más te causo dificultas al resolver las tareas de la guía?
2. ¿Por qué crees que te causó dificultad?
3. ¿Qué fue lo que te pareció más fácil en la guía?
4. Con tus palabras escribe qué aprendiste

¿Qué crees que puedes hacer en la próxima guía para que entiendas mejor lo que se te propone?

Cibergrafía

Referencias

Educación, M. d. (2008). Contenidos para aprender.
MIneducación. (2015). *Vamos a aprender matematicas 11*. Bogota: Graphics.

Anexos

Solución problema 1

SOLUCIÓN

- La media de los datos será \$11.500, es decir, en promedio un estudiante de primer semestre gasta \$11.500 diarios.

  Es importante anotar que el dato \$50.000 está significativamente alejado de la media; por lo tanto, si se calcula el promedio sin incluirlo, este valor cambia notoriamente.

- Como se tiene un número par de datos, la mediana será el promedio de los datos 5 y 6 una vez se hayan ordenado. Por lo tanto:

$$\widetilde{X} = \frac{7.500 + 8.000}{2} = \$7.750$$

De la mediana se puede decir que el 50% de los estudiantes gastan \$7.750 o menos.

La moda es \$8.000; sin embargo, no se puede afirmar que está de moda gastar esta cantidad de dinero en un día regular de universidad.

Solución problema 2

a. Los promedios de llenado de cada una de las maquinas son:

Máquina 1: 352 ml Máquina 2: 351,55 ml

La máquina 2 tiene un llenado medio más cercano al requisito de cumplimiento.

b. Las desviaciones estándar de cada una de las máquinas son

Máquina 1

$$S = \frac{\sqrt{(365 - 352)^2 + \ldots + (352 - 352)^2}}{19} = 10{,}91$$

Máquina 2

$$S = \frac{\sqrt{(380 - 351{,}55)^2 + \ldots + (352 - 351{,}55)^2}}{19} = 35{,}134$$

A pesar de que su promedio es mayor al requerido, la máquina 1 presenta mayor uniformidad en el llenado.