GUIA DE AUTOAPRENDIZAJE Nº12 MATEMÁTICA

IIIº MEDIO

Nombre_______________________________________ Curso:_______ Fecha: _______

Objetivo de Aprendizaje:

OA2: Tomar decisiones en situaciones de incerteza que involucren el análisis de datos estadísticos con medidas de dispersión y probabilidades condicionales.

Objetivo de la guía:

Analizar los datos de situaciones usando medidas de dispersión y tomar decisiones a partir de ello.

Instrucciones: Esta guía es un recurso de acompañamiento y ejercitación de la clase que veras en el video correspondiente, por lo que puedes imprimirla, una vez resuelta y revisada archivarla en una carpeta por asignatura. En caso de no poder imprimir, no hay ningún problema, ya que puedes ir copiando solo los ejemplos en tu cuaderno y dando respuesta a la ejercitación escribiendo el número de pregunta y su respuesta, especificando N° de guía, y fecha**. No olvides que frente a cualquier duda o consulta con respecto a tu clase y/o ejercitación debes contactarnos al correo : matematica.iii.smm@gmail.com**

El video correspondiente a esta clase se encuentra en el link: https://youtu.be/LE0td6bE-y0

Recuerda que **la media aritmética o promedio de un conjunto de datos** $\{x_1, x_2, ..., x_n\}$ se calcula:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

Las medidas de dispersión sirven para **determinar si los datos se encuentran en torno a la media o si están muy dispersos**. Para cuantificar la dispersión, estudiaremos las medidas más conocidas: **el rango, la desviación media, la varianza y la desviación estándar:**

El rango (R) corresponde a la **diferencia entre el mayor y el menor de los datos de la distribución**.

$R = X_{máx} - X_{mín}$

El siguiente **ejemplo** lo utilizaremos para el cálculo de algunas medidas de dispersión:

El entrenador de un equipo de natación debe elegir su representante para la próxima competencia de 100 m en estilo libre. Para ello, cuenta con información consistente en el tiempo, en segundos, de las dos postulantes en las 5 últimas carreras en este estilo:

Competencias de Daniela	
N.° de carrera	Tiempo (s)
1	64
2	58
3	68
4	62
5	65

Competencias de Bárbara	
N.° de carrera	Tiempo (s)
1	69
2	63
3	65
4	50
5	70

La primera medida de dispersión utilizada por el entrenador es el **Rango:**

Si se denotan por R_1 y R_2 los rangos de los tiempos de Daniela y Bárbara respectivamente, se tiene:

$R_1 = X_{máx} - X_{mín} = 68 - 58 = 10 \rightarrow R_1 = 10$ s

$R_2 = X_{máx} - X_{mín} = 70 - 50 = 20 \rightarrow R_2 = 20$ s

Esto da indicios de que los tiempos de Daniela pueden ser menos dispersos que los de Bárbara. Sin embargo, no es posible concluir de inmediato: <u>debemos disponer de más información.</u>

Para disponer de más información, haremos el calculo de otra medida de dispersión **la desviación media:**

- La desviación de una variable x con respecto a su media aritmética está dada por $D = x_i - \bar{x}$.
- La desviación media $(D_{\bar{x}})$ corresponde a la media aritmética de los valores absolutos de las desviaciones $(x_i - \bar{x})$ de los n datos, esto es:

Para datos no agrupados se tiene:

$$D_{\bar{x}} = \frac{|x_1 - \bar{x}| + |x_2 - \bar{x}| + |x_3 - \bar{x}| + \ldots + |x_n - \bar{x}|}{n}$$

Para datos agrupados se tiene:

$$D_{\bar{x}} = \frac{|x_{mc1} - \bar{x}| \cdot f_1 + |x_{mc2} - \bar{x}| \cdot f_2 + |x_{mc3} - \bar{x}| \cdot f_3 + \ldots + |x_{mcn} - \bar{x}| \cdot f_n}{n}$$

Donde x_{mci} es la marca de clase del intervalo i, $\bar{x}$ es la media aritmética de la variable, f_i es la frecuencia absoluta del intervalo i y n es el número total de datos.

La utilización del valor absoluto es porque lo que estamos midiendo es **la distancia del dato con respecto al promedio,** y como bien saben "no existen distancias negativas" y en la resta se podría dar el caso de un resultado negativo.

La **desviación media** permite determinar en cuánto varían, en promedio, los datos de una distribución con respecto a la media aritmética. Entonces calculemos las desviaciones medias de las competidoras:

Como primer paso se calcula el promedio de cada competidora:

$$\overline{X_{Daniela}} = \frac{64+58+68+62+65}{5} = 63{,}4$$
$$\overline{X_{Barbara}} = \frac{69+63+65+50+70}{5} = 63{,}4$$

Se observa que ambas competidoras tienen el mismo promedio en los tiempos.
El segundo paso será hacer el calculo de **las diferencias entre el dato y el promedio, luego sumarlas y dividirlas por el número total de datos,** es decir, la desviación media:

$$\boldsymbol{D_{\overline{Daniela}}} = \frac{\mathbf{|64-63{,}4|+|58-63{,}4|+|68-63{,}4|+|62-63{,}4|+|65-63{,}4|}}{\mathbf{5}} = \mathbf{2{,}72}\ \boldsymbol{s}$$

$$\boldsymbol{D_{\overline{Barbara}}} = \frac{\mathbf{|69-63{,}4|+|63-63{,}4|+|65-63{,}4|+|50-63{,}4|+|70-63{,}4|}}{\mathbf{5}} =$$
$$= \frac{5{,}6+3{,}4+1{,}6+13{,}4+6{,}6}{5} = \mathbf{6{,}12}\ \boldsymbol{s}$$

Concluye, según los resultados, ¿qué datos son más dispersos: los de Daniela o los de Barbara?, ¿por qué?

El entrenador continúa su análisis para tomar una adecuada decisión. Para ello, utilizará otras medidas de dispersión, **la Varianza** ($\boldsymbol{\sigma^2}$) **y la Desviación Estándar** ($\sqrt{\boldsymbol{\sigma^2}}$)**:**

La varianza y la desviación estándar permiten cuantificar la dispersión dada por la desviación media.

•• **La varianza** ($\boldsymbol{\sigma^2}$) corresponde a la media aritmética de los cuadrados de las desviaciones de los n datos. Se expresa en unidades cuadradas.

•• **La desviación estándar (σ)** se obtiene extrayendo la raíz cuadrada de la varianza. Se expresa en la misma unidad que la variable, por lo que nos puede dar una idea más cercana de lo disperso que es el conjunto.

Para datos no agrupados se tiene:

$$\sigma^2 = \frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+(x_3-\bar{x})^2+...+(x_n-\bar{x})^2}{n}$$

Para datos agrupados se tiene:

$$\sigma^2 = \frac{(x_{mc1}-\bar{x})^2 \cdot f_1+(x_{mc2}-\bar{x})^2 \cdot f_2+(x_{mc3}-\bar{x})^2 \cdot f_3+...+(x_{mcn}-\bar{x})^2 \cdot f_n}{n}$$

Donde x_{mci} es la marca de clase del intervalo i, $\bar{x}$ es la media aritmética de la variable, f_i es la frecuencia absoluta del intervalo i y n es el número total de datos.

Calculemos la varianza y desviación estándar para Daniela y Barbara:

Daniela:

$$\boldsymbol{\sigma^2} = \frac{\mathbf{(64-63{,}4)^2+(58-63{,}4)^2+(68-63{,}4)^2+(62-63{,}4)^2+(65-63{,}4)^2}}{\mathbf{5}}$$

$$= \frac{0.36+29.6+21.6+1.96+2.56}{5} = \frac{55.2}{5} = 11{,}04s^2$$

$$\boldsymbol{\sigma} = \sqrt{11.04s^2} \approx 3{,}32\ s$$

Barbara:

$$\sigma^2 = \frac{\mathbf{(69-63,4)^2+(63-63,4)^2+(65-63,4)^2+(50-63,4)^2+(70-63,4)^2}}{\mathbf{5}}$$

$$= \frac{31.36+11.56+2.56+179.56+43{,}56}{5} = \frac{268.6}{5} = 53.72\ s^2$$

$$\boldsymbol{\sigma} = \sqrt{53.72s^2} \approx 7{,}33\ s$$

Finalmente, si tomamos en cuenta la frase: **"A mayor dispersión, mayor valor de la varianza; a menor dispersión, menor valor de la varianza"** responde:

¿Qué decisión debe tomar el entrenador?, ¿quién debería participar en la próxima competencia: Daniela o Bárbara?

Ejercita:

1. Las temperaturas (en grados Celsius) durante dos semanas en Talca fueron las siguientes:

Temperatura semana 1 (°C)	30	31	30	25	21	20	22
Temperatura semana 2 (°C)	30	29	29	27	26	20	27

Calcula e interpreta las medidas de dispersión (Rango, Desviación media, Varianza y Desviación Estándar).

2. La cantidad de cheques cobrados diariamente en todas las sucursales de un banco el mes anterior se registran en la siguiente tabla:

Cantidad de cheques	Frecuencia
[0, 200[	12
[200, 400[	15
[400, 600[	20
[600, 800[	45
[800, 1000]	21

El jefe de sucursales indica la siguiente preocupación:

"Una **desviación estándar** superior a 200 cheques diarios ocasionará problemas de organización y logística en las sucursales".

¿Deberá preocuparse el jefe de operaciones del banco por la cantidad de empleados que se necesitará el mes siguiente?, ¿qué decidirá?

Recuerda que:

Los intervalos son llamados **clases.**

La **marca de clase x_i** de un intervalo, es el promedio entre el limite superior e inferior del intervalo.

El calculo de la **media o promedio** en datos agrupados (en intervalos), es la suma de las multiplicaciones entre la **marca de clase (x_i) y su respectiva frecuencia absoluta (f_i), dividida por el total de datos (n):**

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

3. La chef de un restaurante acaba de recibir un encargo de barras de chocolate de su proveedor, pero aún no los acepta. Los gramos de cada barra se muestran en el recuadro.

178,60	204,12	206,95
221,13	192,78	209,79
226,80	209,79	215,46
229,63	215,46	218,30

Indicando:

¿Qué decisión tomará la chef?, ¿por qué? Argumenta.

Para concluir

a. ¿Qué significa, con respecto a tu rendimiento académico, que tus notas tengan

una dispersión muy alta? Explica.

b. ¿Por qué es importante determinar la dispersión de un conjunto de datos?

c. ¿Qué fue lo que más te costó aprender en este tema?, ¿y lo más fácil?

Evalúate: Toma de decisiones aplicando medidas de dispersión de datos

1. Calcula el rango, la varianza y la desviación estándar de los siguientes conjuntos de datos:

a) 20, 5, 8, 20, 11

b) 6, 2, 13, 1, 12

2. David, el profesor de Historia, tiene la siguiente información respecto de las notas de su curso en una prueba:

Notas del Tercero A en una prueba de Historia	
Nota	**Frecuencia**
[1,0; 2,0[	4
[2,0; 3,0[	8
[3,0; 4,0[	9
[4,0; 5,0[	11
[5,0; 6,0[	7
[6,0; 7,0]	6

Calcula el promedio y la varianza de los datos.

3. El análisis de la sangre de una persona durante 7 semanas arroja las siguientes cantidades de leucocitos, también llamados glóbulos blancos:

Semana 1 → $3500/mm^3$

Semana 2 → $12\,000/mm^3$

Semana 3 → $4800/mm^3$

Semana 4 → $4100/mm^3$

Semana 5 → $3700/mm^3$

Semana 6 → $6200/mm^3$

Semana 7 → $3100/mm^3$

a. Calcula el rango, la desviación media, la varianza y la desviación estándar.

b. El médico que trata al paciente debe cambiar el tratamiento si el promedio de la cantidad de leucocitos es inferior a $4500/mm^3$y la desviación estándar es inferior a 2000 mm^3. ¿Qué decisión tomará el doctor?, ¿por qué?