MEDIDAS DE DISPERSIÓN

NOMBRE ESTUDIANTE:	CURSO: 3°	TIEMPO DE ESTUDIO Y DESARROLLO: 06/abril – 09/abril

OBJETIVO DE APRENDIZAJE	OBJETIVO DE LA CLASE
Resolver problemas que involucren los conceptos de desviación estándar, varianza, coeficiente de variación, tanto de forma manuscrita como haciendo uso de herramientas tecnológicas digitales.	Resolver problemas que involucran medidas de dispersión en diversos contextos y toma de decisión respecto de ellos.

INSTRUCCIONES GENERALES

- En la siguiente guía se presenta el contenido, ejercicios resueltos y ejercicios propuestos de "Medidas de Dispersión".
- De dicha guía, debe estudiar y desarrollarla en su cuaderno **(NO enviar desarrollo al correo del profesor(a) de matemática)**.
- Las dudas que surjan, puede consultarlas por medio del correo oficial del curso, al profesor Manuel Cantillana Matus (manuel.cantillana@liceooscarcastro.cl).
- El tiempo que debe destinar para el estudio y desarrollo será desde el 06 de abril hasta el 09 de abril de 2020.
- La **retroalimentación** de la guía se enviará al correo oficial de cada curso, el 09 de abril de 2020.
- Cuando se retorne a clases presenciales, la o el profesor(a) revisaran su cuaderno, con todos los desarrollos, con el fin de **evaluar proceso** con una nota acumulativa.
- Cada **nota de proceso** se promediará y será consignada en el **Libro de Clases**.

NOTA IMPORTANTE: La Guía anterior ("Medidas de Tendencia Central"), también debía desarrollarse en el cuaderno y será considerada como nota de proceso.

Medidas de Dispersión

A diferencia de las medidas de tendencia central, las medidas de dispersión no son datos de la muestra, más bien, corresponden a parámetros estadísticos que indican que tan cerca o lejos, se encuentran los datos respecto se la media, por lo mismo siempre van a depender de ella. En general, las medidas de dispersión sirven como indicadores de la variabilidad de los datos.

1. <u>Rango</u>

Es la medida de dispersión más sencilla y corresponde a la **distancia de los valores extremos de la muestra**. A mayor rango, se estima mayor dispersión de los datos. También se conoce como amplitud o recorrido.

<u>Ejemplo:</u> Las edades de los estudiantes de un curso fluctúan entre los 15 y los 17 años. Luego el rango es de 2 años $(17-15)$, por lo que se estima que los datos están concentrados.

2. <u>Error muestral</u>

El error se entenderá como la **distancia o diferencia (en valor absoluto) entre un dato específico y la media o promedio**. El error es la base para el cálculo de las medidas de dispersión, por ello, es vital para concluir la variabilidad de los datos.

<u>Ejemplo:</u> Las notas de Andrés en la asignatura de matemática fueron de un 62, 64, 65 y 65. Obteniendo un promedio de 64. Así el error de la primero nota es de 2 puntos $|64-62|=2$, la segunda nota no posee error $|64-64|=0$ y las dos últimas tienen un error de un punto $|65-64|=1$.

3. Desviación Media

La desviación media es un parámetro de dispersión que presenta el **promedio de los errores muestrales**. Si bien es un parámetro considerable, es la menos robusta de las medidas de dispersión.

Ejemplo: Las edades de un grupo de cuatro mamás que participan en un baile para bebé, son 28, 31, 34 y 35. Su desviación media depende del promedio que es 32:

$$DM = \frac{|32-28| + |32-31| + |34-32| + |35-32|}{4} = \frac{4+1+2+3}{4} = \frac{10}{4} = 2.5$$

4. Desviación Típica o Estándar

La desviación típica (también conocida como desviación estándar) es una medida que se usa para cuantificar la variación o dispersión de un conjunto de datos numéricos. Presentando un número cual al ser interpretado expondrá que tan cercanos o lejanos están los datos.

Una desviación estándar baja indica que la mayor parte de los datos de una muestra tienden a estar agrupados cerca de su promedio, mientras que una desviación estándar alta indica que los datos se extienden sobre un rango de valores más amplio. El concepto de alta o baja varía según cada muestra.

Si los datos son 700, 701, 701 y 702 su desviación estándar (σ) será 0,7 $aprox$. Mientras que si los datos son 0,1, 0,4, 0,6 y 1,1 su desviación es de 0,46 $aprox$ y es "más alta" que la anterior. Por lo mismo los valores siempre deben quedar sujeto a interpretación.

Principalmente la desviación estándar permite comparar la **homogeneidad de dos muestras diferentes**, es decir, el nivel de "cercanía" entre los datos de cada muestra.

Se calcula con la fórmula:

$$\sigma = \sqrt{\frac{\sum_i^N (X_i - \bar{X})^2}{N}}$$

Donde

N número total de datos (tamaño de la muestra).
$\bar{X}$ es la media aritmética de la muestra.
X_i es cada valor de la muestra.

Ejemplo: Se tiene la muestra con los datos 700, 701, 701 y 702. Así como su $\bar{X} = 701$, su desviación será:

$$\sigma = \sqrt{\frac{(700-701)^2 + (701-701)^2 + (701-701)^2 + (702-701)^2}{4}}$$

$$= \sqrt{\frac{1^2 + 0^2 + 0^2 + 1^2}{4}} = \sqrt{\frac{2}{4}} = \sqrt{\frac{1}{2}} \approx 0{,}707..$$

5. Varianza

La varianza corresponde a la media de las desviaciones cuadráticas de una variable de carácter aleatorio, y representa la variabilidad de una serie de datos respecto a su media. **La varianza corresponde al cuadrado de la desviación estándar**, sin embargo, y a diferencia de ella, queda expresada en unidades cuadráticas, por lo cual no siempre es interpretable.

La varianza sigue la fórmula descrita por

$$VAR(X) = \sigma^2 = \frac{\sum_i^N (X_i - \bar{X})^2}{N}$$

Para el estudio de distribuciones normales, la varianza corresponde a uno de los parámetros más relevantes, por lo cual, a pesar de no ser tan útil como medida de dispersión, si lo es en el estudio de tipos de distribuciones. Estos contenidos serán trabajados más adelante en el curso.

Ejemplo: Calcular la varianza de las alturas en cm de tres personas que miden: $165\ cm$, $172\ cm$ y $176\ cm$. Sabiendo que la media entre las alturas es de $171\ cm$, se tiene que:

$$\sigma^2 = \frac{(165-171)^2 + (172-171)^2 + (176-171)^2}{3} = \frac{6^2 + 1^2 + 5^2}{3}$$

$$= \frac{36+1+25}{3} = \frac{62}{3} \approx 20{,}66 \dots cm^2$$

Luego la varianza corresponde a $20{,}66\ cm^2$. Lo cual no posee una interpretación relacionada con el ejercicio, esto por la cuadratura de las unidades de medidas.

6. Coeficiente de Variación

El coeficiente de variación es la relación entre desviación estándar de una muestra con su respectiva media aritmética (promedio), y permite principalmente **comparar la homogeneidad de dos muestras diferentes que poseen la misma variable**, pero en diferentes unidades. A menor coeficiente de variación, más homogénea será la muestra (los datos tienden a *acercarse* a la media).

Sigue la fórmula:

$$c.v. = \frac{\sigma}{\bar{X}}$$

Ejemplo: Se comparan dos muestras, una con los valores 700, 701, 701 y 702. Mientras que la segunda muestra tiene valores 0,1, 0,4, 0,6 y 1,1. Así:

Muestra	X_1	X_2	X_3	X_4	$\bar{X}$	σ	$c.v.$
1	700	701	701	702	701	0,7	0,00099
2	0,1	0,4	0,6	1,1	0,55	0,36	0,65454

El coeficiente de variación explicita que, aunque la desviación es mucho mayor en la muestra 1, los datos son mucho más homogéneos que en la muestra 2.

7. Ejercicios

01

Se tienen los siguientes datos de una variable X.

10, 12, 14, 16

Respecto de los estadígrafos de X se afirma que:

I) Mediana (X) = 13
II) Varianza (X) = 5
III) Rango (X) = 6

Es(son) verdadera(s):
A) Solo I
B) Solo I y II
C) Solo II y III
D) Solo I y III
E) I, II y III

02

Si a, b y c son tres números enteros cuya desviación estándar es σ, entonces la desviación estándar de $n+a,\ n+b,\ n+c$ con n un número entero positivo, es:

A) $n^2\sigma$
B) σ
C) $\sqrt{n}\sigma$
D) $n\sigma$
E) $2n\sigma$

03

Se tienen los siguientes valores de una variable X:

1	1	5	9

¿Cuál de los siguientes estadísticos de X es **Falso**?

A) La mediana es 3
B) La media aritmética es 4
C) El rango es 8
D) La varianza es 11
E) La desviación estándar es 8

04 ¿Cuál(es) de las siguientes afirmaciones es (son) siempre verdadera(s)?

I) Si todos los datos numéricos de una población son iguales, entonces la desviación estándar de esta población es 0.
II) Si dos poblaciones de datos numéricos tienen igual promedio, entonces sus varianzas son iguales.
III) Si todos los datos de una población son aumentados en k, con k un entero positivo, entonces su varianza no se altera.

A) Sólo I
B) Sólo III
C) Sólo II y III
D) Sólo I y III
E) I , II y III

05 Si las edades en años, de una población de 8 niños son 2, 3, 5, 6, 8, 10, 11 y 19, entonces su desviación estándar, en años es:

A) $\sqrt{26}$
B) $\sqrt{13}$
C) $\frac{\sqrt{13}}{2}$
D) $\frac{\sqrt{26}}{2}$
E) Ninguna de las anteriores

06 Se tiene una muestra de datos n_1, n_2, n_3 y n_4, donde μ es el promedio. Si a la muestra se le agrega un dato p. ¿Cuál de la siguientes afirmaciones es o son verdaderas?

I) Si $p = \mu$ la desviación estándar aumenta.
II) Si $p = 0$ la desviación estándar disminuye.
III) Si n_1, n_2, n_3, n_4 y p son enteros consecutivos, la desviación estándar es $\sqrt{2}$.

A) Solo I
B) Solo II
C) Solo III
D) Solo I y III
E) Solo II y III

07 ¿Cuál de las siguientes alternativas es **FALSA**?

A) Una desviación estándar pequeña, significa que los datos están concentrados cerca de la media aritmético.
B) Una desviación estándar grande, indica poca confianza en la media aritmética.
C) La desviación estándar siempre es no negativa.
D) Dos muestras con igual número de datos y con la misma media aritmética, tienen desviaciones estándar iguales.
E) La desviación estándar siempre se mide en la misma unidad que los datos.

08

Al analizar los puntajes de los 4 controles realizados por Juan y Pedro, se tuvieron los siguientes resultados.

	Juan	Pedro
Promedio	613	613
Desviación estándar	54,47	168,74

De acuerdo con esta información, ¿Cuál (es) de las siguientes afirmaciones es (son) siempre verdadera (s)?

I) Juan tiene puntajes más cercanos a su promedio.
II) Ambos han obtenido los mismos puntajes en los controles.
III) Existe un error en el cálculo de las desviaciones estándar de Pedro o de Juan, porque ambos tienen el mismo promedio.

A) Solo I
B) Solo I y II
C) Solo I y III
D) Solo II y III
E) I, II y III

09

Si todos los datos de una muestra se multiplican por 4, ¿Cuál(es) de las siguientes afirmaciones es (son) verdadera (s)?

I) El promedio se cuadruplica.
II) La desviación típica se cuadruplica.
III) La varianza se duplica.

A) Solo I
B) Solo II
C) Solo I y II
D) Solo I y III
E) I, II y III

10

Si todos los datos de una muestra se incrementan en 4 unidades, entonces la varianza:

A) Se incrementa en 4 unidades
B) Se incrementa en 2 unidades
C) Queda igual
D) Se incrementa en un 25%
E) Se incrementa en un 50%

11 De acuerdo a la tabla adjunta, ¿Cuál(es) de las siguientes proposiciones es (son) verdadera(s)?

I) $A + B = 3$
II) La desviación estándar es $\sqrt{2}$.
III) La varianza es 2.

x_i	$(x_i - \bar{x})^2$
4	B
5	1
6	0
7	A
8	4

A) Solo I
B) Solo II
C) Solo II y III
D) I, II y III
E) Ninguna de las anteriores

12 Debido a los malos resultados de la prueba de Matemática el profesor decide subir las notas en dos décimas. ¿Cuál de los siguientes estadígrafos **no** cambia?

I) Media
II) Rango
III) Varianza

A) Solo I
B) Solo II
C) Solo III
D) Solo II y III
E) I, II y III

13 ¿Cuál (es) de las siguientes proposiciones es (son) verdadera (s)?

I) La desviación estándar es un número real no negativo.
II) La diferencia entre un dato y el promedio de la muestra puede ser negativa.
III) El rango es una medida de dispersión que puede ser negativa.

A) Solo I
B) Solo I y II
C) Solo II y III
D) I, II y III
E) Ninguna de ellas

14 La desviación estándar de los datos $4a$, $4b$ y $4c$ es 0,16, entonces la desviación estándar de los datos a, b y c es igual a:

A) 0,1
B) 0,04
C) 0,16
D) 0,64
E) 1

15

En un supermercado todo los fines de semana los artículos están rebajados en un 10%, si se considera una muestra de 100 artículos, entonces ¿Cuál(es) de los siguientes estadígrafos de la muestra también variarían en el mismo porcentaje?

I) Media
II) Rango
III) Desviación estándar

A) Solo I
B) Solo II
C) Solo III
D) Solo I y III
E) I, II y III

16

Una empresa posee sucursales en Chile como el Portugal, a un gran porcentaje de los trabajadores se les paga el sueldo mínimo de cada país ($\$350.000\ y\ 700€$) y se presentan como datos la media de los sueldos y la desviación estándar de esta última.

	Media	Desviación E.
Chile	$414.285	$123098
Portugal	800€	165€

Con los datos presentados, exprese opinión respecto a la comparación entre los sueldos pagados por la misma empresa en ambos países.