

# Apuntes Unidad 2

Media, desviación media absoluta, varianza y  
desviación estándar

---

Curso: Probabilidad y estadística descriptiva

Unidad 2 : Media muestral, dispersión y correlación

Tema: Medidas de tendencia central y de dispersión

Contenido: Media, desviación media absoluta, varianza y desviación estándar

## MEDIA MUESTRAL

La **media muestral**, que se denota  $\bar{x}$ , corresponde al promedio de los datos de una muestra. Tomemos el ejemplo de una muestra que tiene 5 datos:  $x_1 = 3$ ,  $x_2 = 6$ ,  $x_3 = 9$ ,  $x_4 = 15$  y  $x_5 = 17$ . Su media muestral es:

$$\bar{x} = \frac{3+6+9+15+17}{5} = 10$$

Una intuición que se puede generar al pensar en cómo medir la dispersión de una distribución es comparar cada dato con la media muestral. Para realizar esto, podemos restar cada dato con la media y anotar el resultado.

Operación	Resultado
$3 - 10 =$	-7
$6 - 10 =$	-4
$9 - 10 =$	-1
$15 - 10 =$	5
$17 - 10 =$	7

Ahora, como queremos estudiar la dispersión de todos los datos, procedemos a sumar los resultados.

$$-7 + (-4) + (-1) + 5 + 7 = 0$$

Sorprendentemente da cero. Esto ocurre ya que la media cumple esa propiedad, independiente de cuales sean los datos de la muestra, al sumar las diferencias entre los datos y la media siempre dará cero. Considerando que la intuición inicial es correcta, utilizaremos la distancia de los datos con respecto a la media muestral, cálculo conocido como **desviación media absoluta**.

## DESVIACIÓN MEDIA ABSOLUTA

La **desviación media absoluta** es una medida de dispersión que se define como el promedio de las distancias de los datos a la media. Por ejemplo, si tenemos un conjunto de cinco datos, 4, 5, 9, 15 y 17, cuya media es 10, se tiene que la desviación media absoluta es:

Curso: Probabilidad y estadística descriptiva

Unidad 2 : Media muestral, dispersión y correlación

Tema: Medidas de tendencia central y de dispersión

Contenido: Media, desviación media absoluta, varianza y desviación estándar

$$\frac{|4-10| + |5-10| + |9-10| + |15-10| + |17-10|}{5} = \frac{24}{5} = 4,8$$

Con esta nueva medida, podremos medir la dispersión de distintas distribuciones de datos con respecto a la media de la muestra.

## VARIANZA MUESTRAL

La **varianza muestral**, que se denota como  $s^2$ , es una medida de la dispersión de los datos de una muestra con respecto a la media muestral  $\bar{x}$ . A diferencia de la desviación media absoluta, se calcula a partir de la suma los cuadrados de las distancias de los datos con respecto a  $\bar{x}$ . Esta cantidad luego se divide por  $(n - 1)$  donde  $n$  la cantidad de datos de la muestra. A continuación, podemos observar la expresión matemática que la define.

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}$$

En algunas ocasiones, como por ejemplo en una muestra con gran cantidad de datos, puede ser engorroso calcular la varianza muestral. Debido a esto, aprenderemos a calcularla, al igual que las medidas vistas anteriormente, en algunos programas computacionales.

## CALCULAR MEDIDAS DE DISPERSIÓN USANDO PROGRAMAS COMPUTACIONALES

A continuación, aprenderemos a calcular las medidas de dispersión que hemos visto hasta el momento (desviación media absoluta y varianza) usando una hoja de cálculo. Esto se puede realizar con Microsoft Excel o con Google Sheets, pero como los procedimientos son análogos y los comandos que usan son los mismos, solo revisaremos el procedimiento para el primer programa.

### Microsoft Excel

**Paso 1:** En primer lugar, debemos calcular la media muestral de los datos. Para ello, usamos el comando "PROMEDIO", el cual recibe dos parámetros que representan el rango donde se encuentra la muestra, es decir, el primer parámetro es la primera celda que contiene datos de la muestra y el segundo es la última celda donde se tienen datos.

**Paso 2:** Para calcular la distancia respecto a la media muestral, usaremos el comando "ABS", que entrega el valor absoluto. Luego, en otra celda se debe escribir el comando utilizando la siguiente idea: =ABS(CD - CMM), donde CD es la celda que contiene el dato y CMM la celda que contiene la media muestral. Es importante mencionar que la celda "CMM" debe tener un signo peso al inicio y otra entre la letra y el número, es decir, si tenemos la celda "A1", que contiene a la media muestral, debemos escribir \$A\$1.

Curso: Probabilidad y estadística descriptiva

Unidad 2 : Media muestral, dispersión y correlación

Tema: Medidas de tendencia central y de dispersión

Contenido: Media, desviación media absoluta, varianza y desviación estándar

**Paso 3:** Ahora debemos arrastrar la celda que contiene la distancia calculada en el paso 2 para todos los datos.

**Paso 4:** Ahora que tenemos la distancia de cada dato respecto a la media, podemos calcular la desviación media absoluta, la que corresponde al promedio de estos valores. Para ello, repetimos el procedimiento usado para calcular la media muestral, pero ahora con las distancias calculadas.

**Paso 5:** Para la varianza necesitamos calcular el cuadrado de las distancias respecto a la media. Ahora, debemos sumar los cuadrados de las distancias y luego dividir este valor por  $(n - 1)$  donde  $n$  la cantidad de datos de la muestra.

En muchas ocasiones la varianza puede ser difícil de interpretar, es por eso que introduciremos una nueva medida de dispersión. A continuación, definiremos la desviación estándar muestral.

## DESVIACIÓN ESTÁNDAR MUESTRAL

La **desviación estándar muestral** de un conjunto de datos, que se denota como  $s$ , corresponde a la raíz cuadrada positiva de la varianza muestral:

$$s = \sqrt{s^2}$$

La desviación estándar muestral tiene la misma unidad de medida que la variable que se está analizando, al ser la raíz cuadrada de esta.

**Nota:** Podemos observar que a medida que se alejan los datos de la media, todas las medidas de dispersión aumentan. Sin embargo, la varianza, y por tanto la desviación estándar, aumentan más que la desviación media absoluta si es que los puntos que se alejan son aquellos que están a una mayor distancia. Esto se debe a que, como la varianza depende del cuadrado de la distancia, se ve más afectada por cambios en los datos que están lejos de la media que en aquellos que están más cerca.

## RELACIÓN ENTRE DESVIACIÓN ESTÁNDAR Y DESVIACIÓN MEDIA ABSOLUTA

La desviación media absoluta, la varianza y la desviación estándar son medidas de dispersión que se usan para describir cuán dispersos están los datos con respecto a la media muestral. En relación a estas medidas de dispersión, se tiene lo siguiente:

- La desviación estándar es siempre mayor o igual que la desviación media absoluta.
- La desviación estándar se ve más afectada que la desviación media absoluta a cambios en los datos alejados de la media que en aquellos que están más cerca de la media.

Curso: Probabilidad y estadística descriptiva

Unidad 2 : Media muestral, dispersión y correlación

Tema: Medidas de tendencia central y de dispersión

Contenido: Media, desviación media absoluta, varianza y desviación estándar

En algunas ocasiones, el cálculo de las medidas de dispersión puede ser algo extenso y complejo, por lo que a continuación aprenderemos cómo calcularlas utilizando Microsoft Excel.

## CALCULAR DESVIACIÓN MEDIA ABSOLUTA, VARIANZA MUESTRAL Y DESVIACIÓN ESTÁNDAR USANDO UNA HOJA DE CÁLCULO

En lecciones anteriores se vieron distintas maneras de calcular estas medidas. Microsoft Excel y Google Sheets utilizan algunos comandos que permiten agilizar el cálculo. A continuación, se pueden observar los comandos necesarios.

Medida de dispersión	Excel 2007 o anterior	Excel posterior a 2007	Google Sheets
Desviación media absoluta	DESVROM()	DESVROM()	DESVROM()
Varianza muestral	VAR()	VAR.S()	VAR()
Desviación estándar muestral	DESVEST()	DESVEST.M()	DESVEST()

Para utilizar estos comandos debes incluir entre el paréntesis la ubicación de los datos. Por ejemplo, para calcular la varianza muestral de una muestra, suponiendo que se tiene Excel 2017 y los datos ubicados en una columna desde la celda A1 hasta A55, se debe utilizar VAR(A1:A55).

## SÍNTESIS

- La media muestral, que se denota  $\bar{x}$ , corresponde al promedio de los datos de una muestra.
- La desviación media absoluta es una medida de la dispersión de los datos de una muestra, que se calcula como el promedio de la distancia a la que están los datos con respecto a la media muestral.
- La varianza muestral es una medida de dispersión de los datos respecto a la media. Para calcularla es necesario sumar **los cuadrados de las distancias** de los datos con respecto a la media muestral y luego dividir el resultado por  $(n - 1)$ .

**Curso:** Probabilidad y estadística descriptiva

**Unidad 2 :** Media muestral, dispersión y correlación

**Tema:** Medidas de tendencia central y de dispersión

**Contenido:** Media, desviación media absoluta, varianza y desviación estándar

- La desviación estándar es otra medida de dispersión que corresponde a la raíz cuadrada de **la varianza muestral**. Su unidad de medida es la misma que la de los datos, lo que hace que sea más fácil interpretar su valor respecto a la varianza muestral.
- La desviación estándar es siempre mayor o igual que la desviación media absoluta.
- La desviación estándar se ve más afectada que la desviación media absoluta a cambios en los datos alejados de la media que en aquellos que están más cerca de la media.