

# Varianza y desviación estándar

*La desviación sólo significa qué tan lejos de lo normal*

## Desviación estándar

La desviación estándar ( $\sigma$ ) mide cuánto se separan los datos.

La fórmula es fácil: es la raíz cuadrada de la **varianza**. Así que, "¿qué es la varianza?"

## Varianza

la varianza (que es el cuadrado de la desviación estándar:  $\sigma^2$ ) se define así:

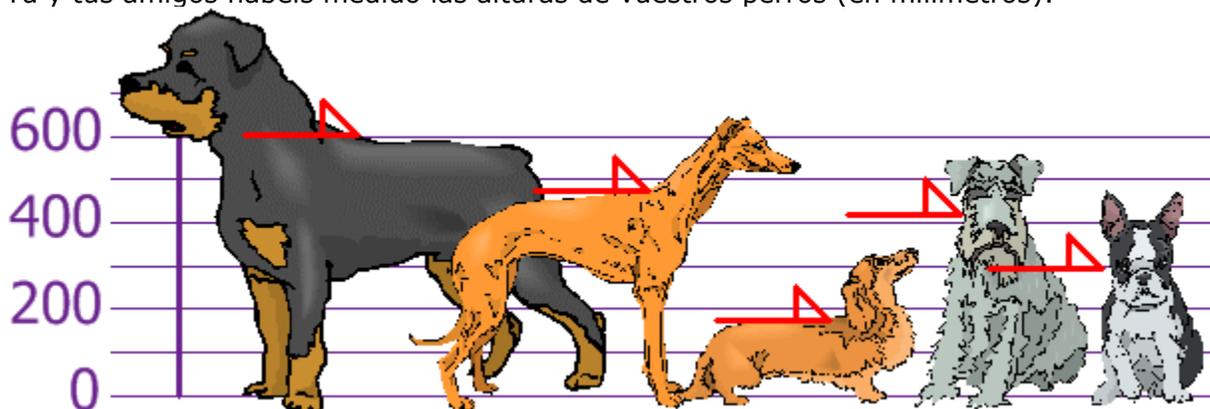
Es la media de las diferencias con la media **elevadas al cuadrado**.

En otras palabras, sigue estos pasos:

1. Calcula la [media](#) (el promedio de los números)
2. Ahora, por cada número resta la media y eleva el resultado al cuadrado (la diferencia elevada al cuadrado).
3. Ahora calcula la media de esas diferencias al cuadrado. ([¿Por qué al cuadrado?](#))

## Ejemplo

Tú y tus amigos habéis medido las alturas de vuestros perros (en milímetros):



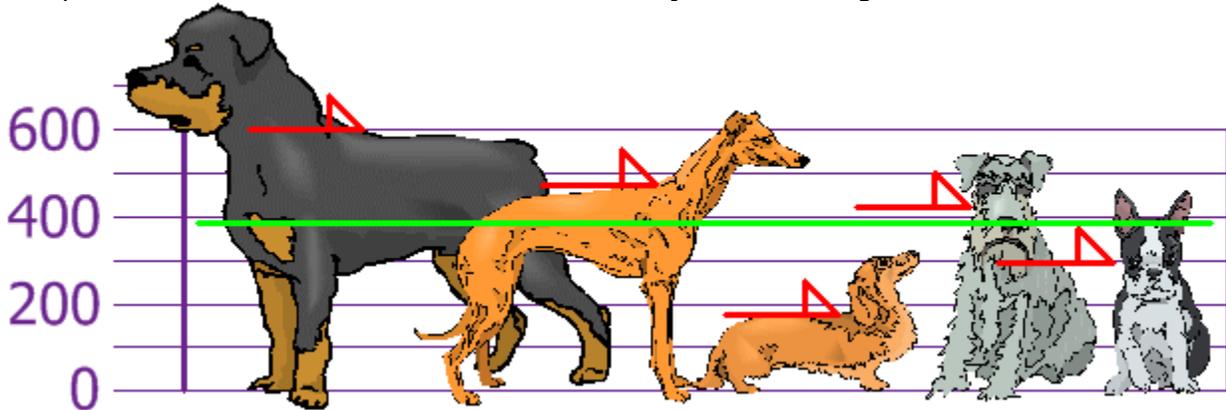
Las alturas (de los hombros) son: 600mm, 470mm, 170mm, 430mm y 300mm.

Calcula la media, la varianza y la desviación estándar.

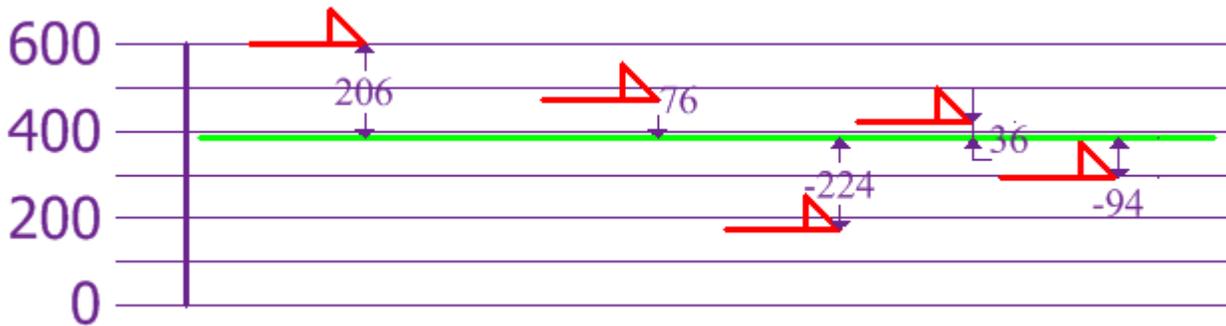
Respuesta:

$$\text{Media} = \frac{600 + 470 + 170 + 430 + 300}{5} = \frac{1970}{5} = 394$$

así que la altura media es 394 mm. Vamos a dibujar esto en el gráfico:



Ahora calculamos la diferencia de cada altura con la media:



Para calcular la varianza, toma cada diferencia, elévala al cuadrado, y haz la media:

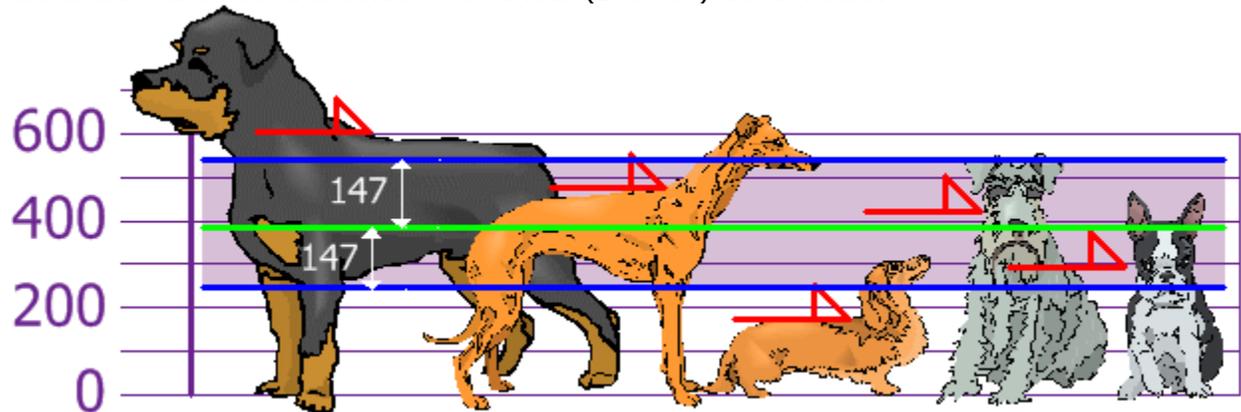
$$\text{Varianza: } \sigma^2 = \frac{206^2 + 76^2 + (-224)^2 + 36^2 + (-94)^2}{5} = \frac{108,520}{5} = 21,704$$

Así que la varianza es 21,704.

Y la desviación estándar es la raíz de la varianza, así que:

$$\text{Desviación estándar: } \sigma = \sqrt{21,704} = 147$$

y lo bueno de la desviación estándar es que es útil: ahora veremos qué alturas están a distancia menos de la desviación estándar (147mm) de la media:



Así que usando la desviación estándar tenemos una manera "estándar" de saber qué es normal, o extra grande o extra pequeño.

Los Rottweilers **son** perros grandes. Y los Dachsunds **son** un poco menudos... ¡pero que no se enteren!

### **\*Nota: ¿por qué al cuadrado?**

Elevar cada diferencia al cuadrado hace que todos los números sean positivos (para evitar que los números negativos reduzcan la varianza)

Y también hacen que las diferencias grandes se destaquen. Por ejemplo  $100^2=10,000$  es mucho más grande que  $50^2=2,500$ .

Pero elevarlas al cuadrado hace que la respuesta sea muy grande, así que lo deshacemos (con la raíz cuadrada) y así la desviación estándar es mucho más útil.