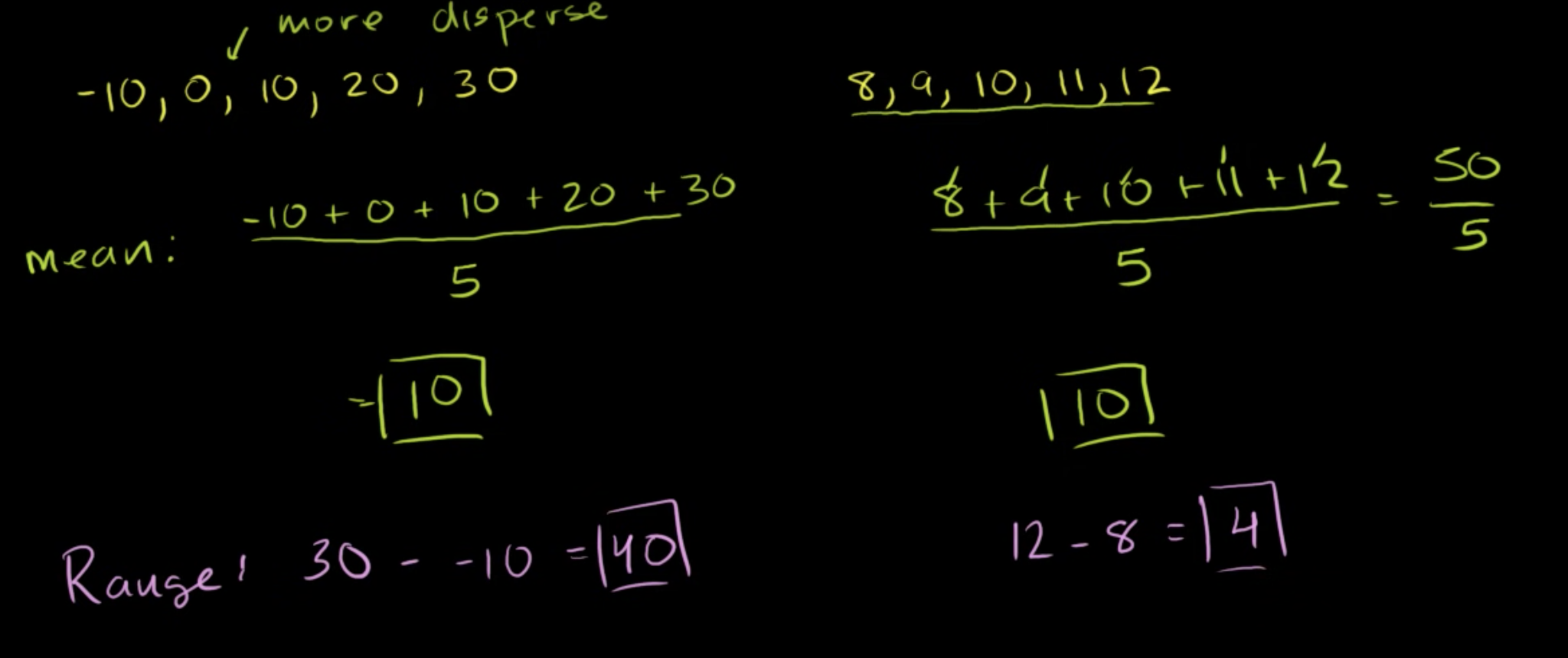
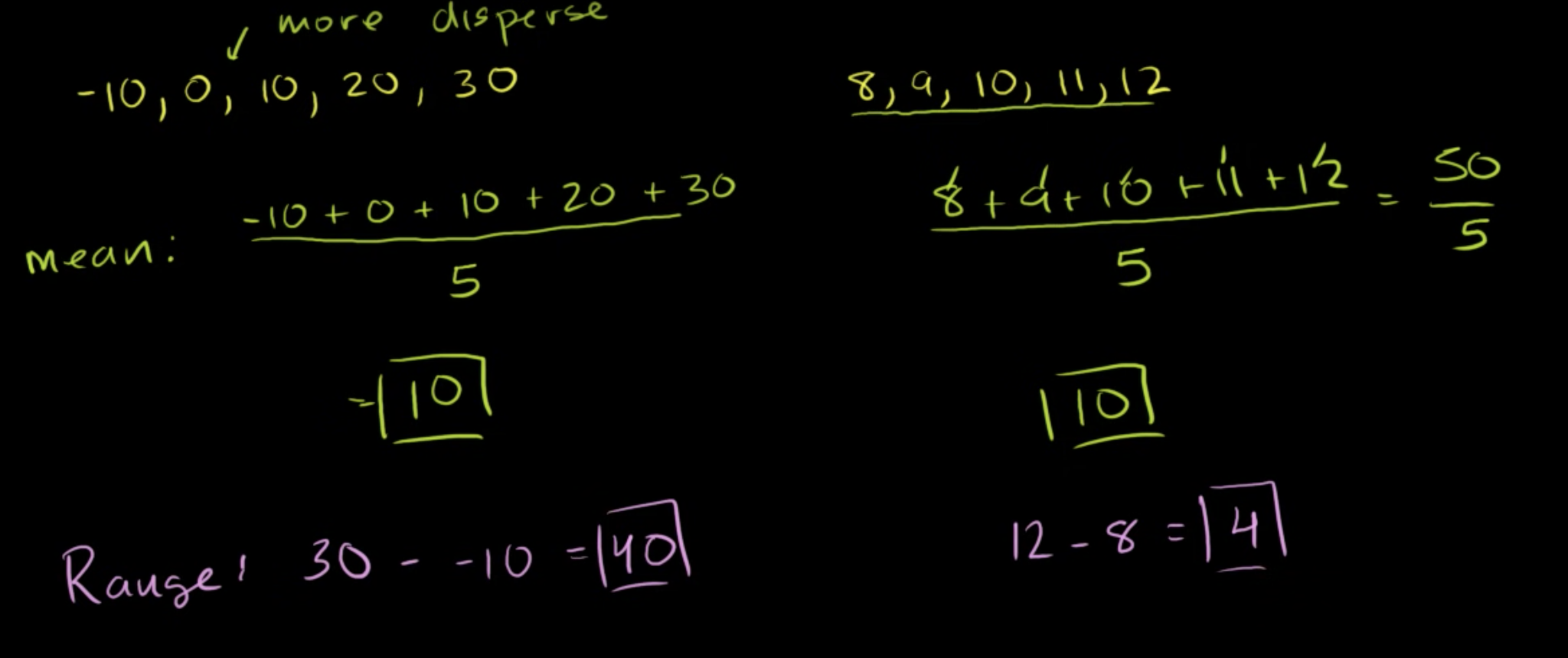
**Medidas de DISPERSIÓN - Introducción a la estadística**

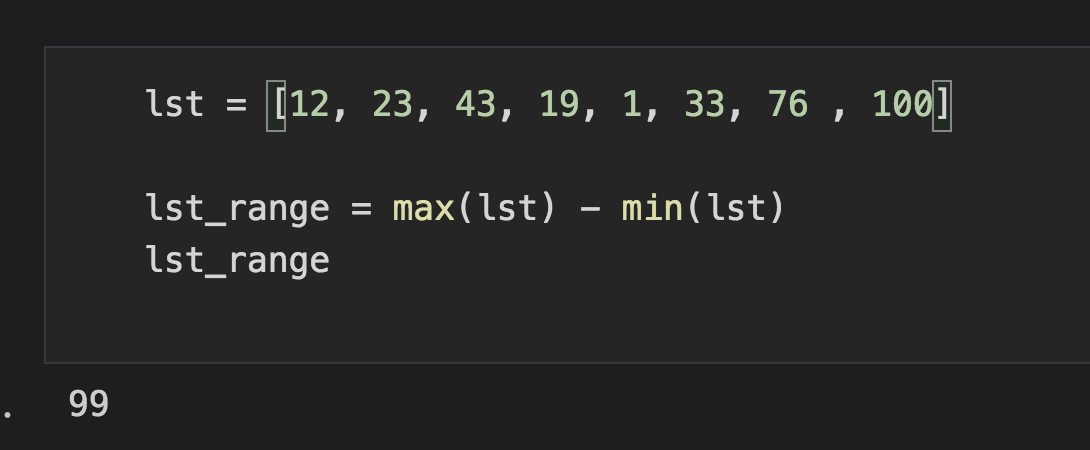
Las medidas de dispersión se utilizan para medir qué tan "dispersos" están los datos. Aquí hablaremos de "Rango", "Standard Deviation" y "Variance".

[video sencillo explicativo](https://www.youtube.com/watch?v=E4HAYd0QnRc)

1. **RANGO:** Es la diferencia entre los dos extremos del conjunto de datos. Si `X max` y `X min` son las dos datos extremos, entonces `Range` =` X max` - `X min`





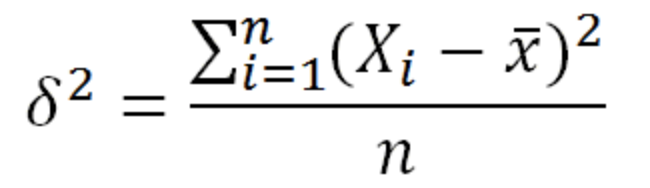


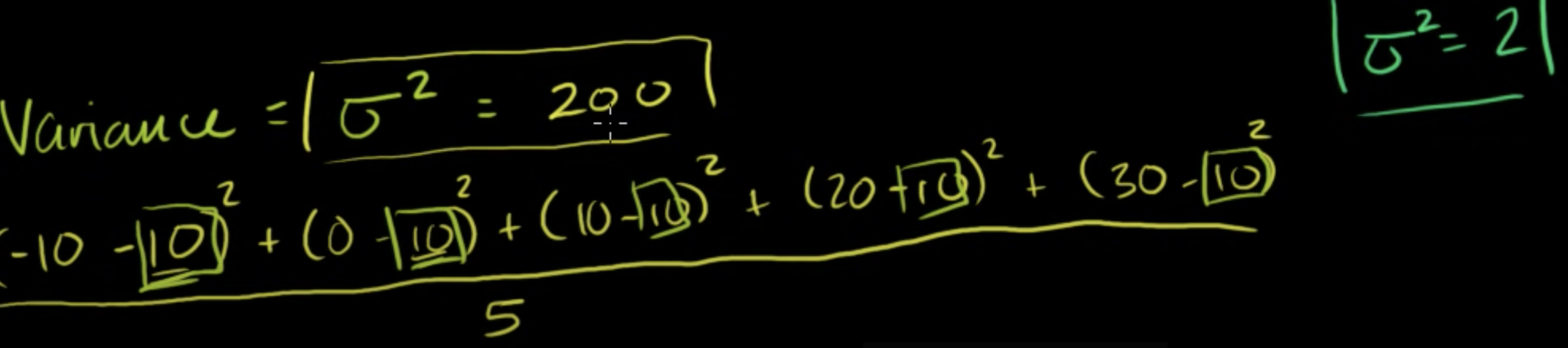
1. **VARIANCE:** La varianza describe la dispersión en los datos respecto a la media (mean). *Varianza = sigma al cuadrado*

→ Esta cantidad siempre es positiva

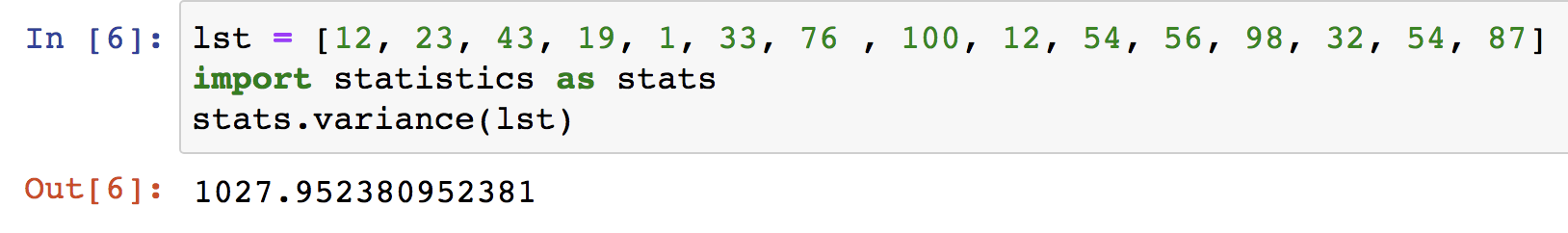
→ Cuanto mayor es la magnitud, mayor es la propagación y viceversa

[(X1-mean)º2 + … + (Xn-mean)º2]/n → (Xn-mean)=deviation

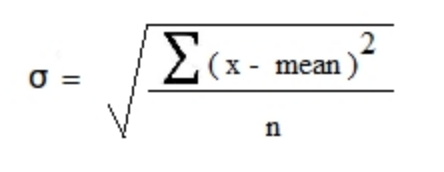




**→ PYTHON: stats.variance()**

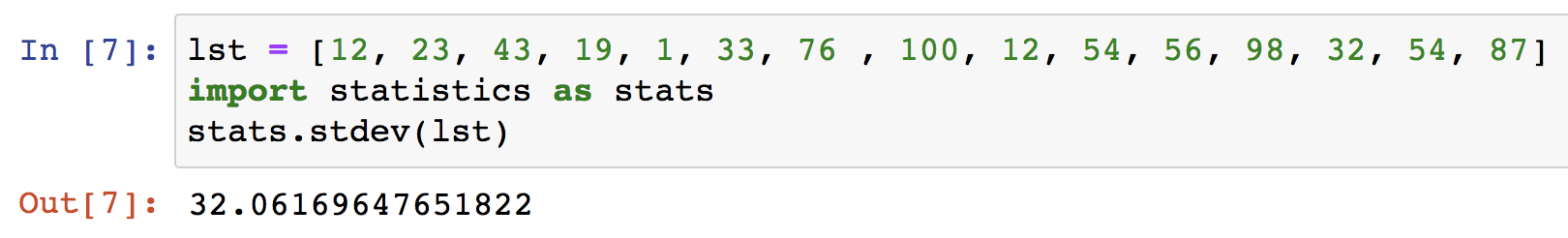


1. **STANDARD DEVIATION:** Es la raíz cuadrada de la varianza.



→ (Xn-mean)=deviation por lo tanto será el sumatorio de las desviaciones entre la raíz cuadrada de “n”

**→ PYTHON: stats.stdev()**

****

DIFERENCIAS ENTRE LA VARIANZA Y LA DESVIACIÓN ESTÁNDAR:

Aunque tanto la varianza como la desviación estándar son medidas de extensión/dispersión, existe una diferencia en las unidades de las dos cosas.

* La unidad de varianza es el cuadrado de la unidad de los datos originales
* La unidad de desviación estándar es la misma que la unidad de los datos originales.

→ Por lo tanto, para fines prácticos, a veces la gente prefiere usar la desviación estándar en lugar de la varianza.

*Además, dado que la varianza es el cuadrado de la desviación estándar, si el valor de la desviación estándar es grande, la magnitud de la varianza aumenta. A veces es preferible trabajar con números de menor magnitud.*

**POPULATION and SAMPLE**:

Cuando hablamos de estadística descriptiva, solo hablamos de los datos como un todo, es decir. A esto se le llama datos de **"POPULATION"**.

Pero en el mundo real no es posible recopilar una cantidad tan grande de puntos de datos. Por lo tanto, trabajamos con **un conjunto más pequeño** (un subet de datos). Este subconjunto también se denomina **“SAMPLE”**. Se supone que esta muestra proviene de una gran población y se toma al azar, por lo que la muestra extraída también se llama **muestra aleatoria**.

→ Siempre que trabajamos con una muestra, las estadísticas descritas anteriormente se denominan “sample statistics”:

* mean → SAMPLE MEAN
* variance → SAMPLE VARIANCE
* standard deviation → SAMPLE STANDARD DEVIATION