Mean is the value you would expect on average.

Variance is the concentration of datapoints around the mean value.

Example: 1D Variance

$$D_1 = \{1, 2, 4, 5\}, \quad E[D_1] = 3.$$
$$D_2 = \{-1, 3, 7\}, \quad E[D_2] = 3$$

$$D_1: \frac{(1-3)^2 + (2-3)^2 + (4-3)^2 + (5-3)^2}{4} = 10/4$$

$$D_2: \frac{(-1-3)^2 + (3-3)^2 + (7-3)^2}{3} = 32/3$$

The average squared deviation of D_2 is greater then D_1 – meaning the spread (less concentrated) of the data points is greater for D_2.

The idea of spread and variance can be formalized.

$X =: \{x_1, \dots, x_n\}$, a dataset

The average squared deviation—that is the variance,

$$Var[X] = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)^2, \quad \text{where } \mu = E[X]$$